



ESTIMANDO REDSHIFTS FOTOMÉTRICOS COM REGULARIZAÇÃO SENSÍVEL AOS ERROS

Raphael Correia de Souza Fialho

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação, do Centro Federal de Educação Tecnológica Celso Suckow da Fonseca, CEFET/RJ, como parte dos requisitos necessários à obtenção do título de mestre.

Orientador(a): Eduardo Bezerra, D.Sc.
Coorientador(a): Ricardo Ogando, D.Sc.

Rio de Janeiro,
Novembro 2020

ESTIMANDO REDSHIFTS FOTOMÉTRICOS COM REGULARIZAÇÃO SENSÍVEL AOS
ERROS

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação, do Centro Federal de Educação Tecnológica Celso Suckow da Fonseca, CEFET/RJ, como parte dos requisitos necessários à obtenção do título de mestre.

Raphael Correia de Souza Fialho

Banca Examinadora:

Presidente, Eduardo Bezerra, D.Sc. (CEFET/RJ) (Orientador(a))

Ricardo Ogando, D.Sc. (ON/MCTIC) (Coorientador(a))

Rafaelli de Carvalho Coutinho, D.Sc.(CEFET/RJ)

Ribamar R. de R. dos Reis, D.Sc. (Instituto de Física - UFRJ)

Ronaldo Ribeiro Goldschmidt, D.Sc. (IME/RJ)

Rio de Janeiro,
Novembro 2020

Ficha catalográfica elaborada pela Biblioteca Central do CEFET/RJ

F439 Fialho, Raphael Correia de Souza
Estimando redshifts fotométricos com regularização sensível
aos erros / Raphael Correia de Souza Fialho — 2021.
126f. + anexo : il. color. , enc.

Dissertação (Mestrado) Centro Federal de Educação
Tecnológica Celso Suckow da Fonseca , 2021.

Bibliografia : f. 120-126

Orientador: Eduardo Bezerra

Coorientador: Ricardo Ogando

1. Aprendizado de máquina . 2. Redes neurais (Computação).
3. Algoritmos. 4. Astronomia - Observações. I. Bezerra, Eduardo
(Orient.). II. Ogando, Ricardo (Coorient.). III. Título.

CDD 006.31

RESUMO

Estimando Redshifts Fotométricos com Regularização Sensível aos Erros

Na Astronomia tem se tornado comum o uso de algoritmos de aprendizagem de máquina durante o processo de captura e análise de eventos astronômicos. Devido à quantidade atual de dados capturados pelos telescópios e antenas em levantamentos astronômicos, esses dados costumam ser armazenados, catalogados e transformados para análises e estudos posteriores. Um tipo particular de análise feita sobre esses dados é a detecção do *redshift* fotométrico, medida que está relacionada ao quão distante um objeto (galáxia ou quasar) se encontra em relação a um determinado ponto de referência. Uma característica relevante dos conjuntos de dados relativos a estudos sobre *redshift* fotométrico é que cada entrada apresenta não apenas as medições realizadas para um determinado objeto, mas também um valor de erro correspondente a cada medição. Nesta dissertação estudamos a construção de modelos de predição para *redshift* fotométrico por meio de algoritmos de aprendizado de máquina. Damos foco ao uso de redes neurais artificiais. Nosso objetivo é investigar de que forma esses modelos se comportam em cenários em que a informação sobre erros das medições são considerados ou ignorados durante a etapa de aprendizado. Em particular, propomos uma técnica de treinamento que almeja tirar proveito dos valores de erro. Realizamos experimentos computacionais comparativos para avaliar a efetividade da técnica proposta.

Palavras-chave: Aprendizado de Máquina; Redes Neurais Artificiais; Redshift Fotométrico

ABSTRACT

Estimating Photometric Redshifts with Error-Sensitive Regularization

In Astronomy it has become common to use machine learning algorithms during the process of capturing and analyzing astronomical events. Due to the current amount of data captured by telescopes and antennas in astronomical surveys, these data are usually stored, cataloged and transformed for further analysis and studies. A particular type of analysis done on these data is the prediction of the photometric redshift, a measure that is related to how far an object (e.g., galaxy or quasar) is in relation to a given reference point. A relevant feature of the datasets used for investigating photometric redshift is that each entry presents not only the measurements made for a given object, but also an error value corresponding to each measurement. In this dissertation we study the construction of prediction models for photometric redshift using machine learning algorithms. We focus on the use of artificial neural networks. Our goal is to investigate how these models behave in scenarios where information about measurement errors is considered or ignored during the learning stage. In particular, we propose a training technique that aims to take advantage of the error values. We performed comparative computational experiments to evaluate the effectiveness of the proposed technique.

Keywords: Machine Learning; Artificial Neural Networks; Photometric Redshift

LISTA DE ILUSTRAÇÕES

Figura 1 –	Classificação de Galáxias	21
Figura 2 –	Espectro da luz	22
Figura 3 –	Espectro de uma estrela	25
Figura 4 –	Ferramenta SkyServer	26
Figura 5 –	Diagrama de Hubble	27
Figura 6 –	Deslocamento para o Vermelho	29
Figura 7 –	Representação de um neurônio artificial	33
Figura 8 –	Esquema de uma rede MLP com $(L + 1)$ camadas.	34
Figura 9 –	Curva do regressor ajustado por redes neurais MLP	35
Figura 10 –	Curva do Regressor de Árvores de Decisão	37
Figura 11 –	Curva de Regressor XGBoost	39
Figura 12 –	Curva de Regressor Isotônico: Exemplo 1	40
Figura 13 –	Curva de Regressor Isotônico: Exemplo 2	41
Figura 14 –	Visão geral do <i>Dropout</i> .	49
Figura 15 –	Gráfico magnitude vs erro	51
Figura 16 –	Ajuste da função $h_j(\cdot)$	52
Figura 18 –	Exemplo numérico da técnica de alteração do vetor de magnitudes.	55
Figura 19 –	Matriz de Correlação do Conjunto SDSS DR16	62
Figura 20 –	Boxplot do Conjunto SDSS DR16	63
Figura 21 –	Matriz de Correlação do Conjunto Limpo do SDSS DR16	64
Figura 22 –	Boxplot das Magnitudes do Conjunto Limpo SDSS DR16	65
Figura 23 –	Boxplot dos Erros do Conjunto Limpo SDSS DR16	65
Figura 24 –	Boxplot dos Erros do Conjunto Filtrado SDSS DR16	66
Figura 25 –	Magnitudes por Erro em Cada Banda da Amostragem do Conjunto SDSS DR16	67
Figura 26 –	Boxplot dos Erros do Conjunto Limpo e Filtrado SDSS DR16	68

Figura 27 –	Magnitudes por Erro em Cada Banda da Amostragem do Conjunto Limpo e Filtrado SDSS DR16	69
Figura 28 –	Magnitudes por Erro em Cada Banda da Amostragem do Conjunto SDSS DR16	71
Figura 29 –	Matriz de Correlação do Conjunto COIN/Teddy	73
Figura 30 –	Boxplot das Magnitudes do Conjunto COIN/Teddy	74
Figura 31 –	Boxplot dos Erros do Conjunto COIN/Teddy	74
Figura 32 –	Magnitudes por erro em cada banda do conjunto COIN/Teddy	75
Figura 33 –	Matriz de Correlação do Conjunto COIN/Happy	77
Figura 34 –	Boxplot das Magnitudes do Conjunto COIN/Happy	78
Figura 35 –	Boxplot dos Erros do Conjunto COIN/Happy	79
Figura 36 –	Magnitudes por Erro em Cada Banda do Conjunto COIN/Happy	80
Figura 37 –	Z-Spec e Z-Phot por modelo do conjunto SDSS-R16-120k	93
Figura 38 –	Z-Spec e Z-Phot por modelo do conjunto COIN/Teddy	94
Figura 39 –	Z-Spec e Z-Phot por modelo do conjunto COIN/Happy	95
Figura 40 –	Resíduos por modelo do conjunto SDSS-R16-120k	96
Figura 41 –	Resíduos por modelo do conjunto COIN/Teddy	97
Figura 42 –	Resíduos por modelo do conjunto COIN/Happy	98
Figura 43 –	Coeficiente R2	106

LISTA DE TABELAS

Tabela 1 – Sistema de magnitudes ugriz	24
Tabela 2 – Magnitudes e erros	25
Tabela 3 – Trabalhos relacionados por funcionalidades	45
Tabela 4 – Trabalhos relacionados por métodos	46
Tabela 5 – Perfil da Máquina do Experimento	57
Tabela 6 – Conjunto de Dados do COIN	58
Tabela 7 – Conjunto de Dados do SDSS	60
Tabela 8 – Comparativo dos Arquivos dos Conjuntos de Dados	60
Tabela 9 – Comparativo dos Arquivos de Treinamento, Validação e Teste do Conjunto COIN/Teddy	73
Tabela 10 – Comparativo dos Arquivos de Treinamento, Validação e Teste do Conjunto COIN/Happy	77
Tabela 11 – Tabela de hiperparâmetros usados no Regressor XGBoost	84
Tabela 12 – Resultados sobre o conjunto SDSS-R16-120k	86
Tabela 13 – Resultados sobre o conjunto COIN/Teddy	87
Tabela 14 – Resultados sobre o conjunto COIN/Happy	88
Tabela 15 – Taxas de <i>outliers</i> em erros padrão	90
Tabela 16 – Taxas de <i>outliers</i> em erros padrão	91
Tabela 17 – Taxas de <i>outliers</i> em erros padrão	92
Tabela 18 – Resultados do Conjunto SDSS-R16-120k	107
Tabela 19 – Resultados sobre o conjunto COIN/Teddy	108
Tabela 20 – Resultados sobre o conjunto COIN/Happy	109
Tabela 21 – Resultados do Conjunto SDSS-R16-120k	111
Tabela 22 – Resultados sobre o conjunto COIN/Teddy	112
Tabela 23 – Resultados sobre o conjunto COIN/Happy	113

Tabela 24 – Resultados do Conjunto SDSS-R16-120k	114
Tabela 25 – Resultados sobre o conjunto COIN/Teddy	115
Tabela 26 – Resultados sobre o conjunto COIN/Happy	116
Tabela 27 – Resultados do Conjunto SDSS-R16-120k	117
Tabela 28 – Resultados sobre o conjunto COIN/Teddy	118
Tabela 29 – Resultados sobre o conjunto COIN/Happy	119

LISTA DE ABREVIATURAS E SIGLAS

AM	Aprendizagem De Máquina
COIN	Programa <i>Cosmostatistics Initiative</i>
RNA	Redes Neurais Artificiais
SDSS	<i>Sloan Digital Sky Survey</i>
SDSS-DR16	<i>Sloan Digital Sky Survey (Data Release 16)</i>
SDSS-DR12	<i>Sloan Digital Sky Survey (Data Release 12)</i>

SUMÁRIO

1	Introdução	14
1.1	Contextualização	14
1.2	Justificativa	16
1.3	Objetivos	17
1.4	Metodologia	17
1.5	Organização dos Capítulos	19
2	Fundamentação Teórica	20
2.1	Conceitos Relevantes da Astronomia	20
2.1.1	Galáxias	20
2.1.2	Sistemas de Magnitudes	23
2.1.3	Lei de Hubble	26
2.1.4	Efeito Doppler	28
2.1.5	<i>Redshift</i> Fotométrico	29
2.1.6	Levantamentos Fotométricos	30
2.2	Métodos de Regressão	31
2.2.1	Redes Neurais MLP	32
2.2.2	Árvores de Decisão	35
2.2.3	Algoritmo XGBoost	37
2.2.4	Regressão Isotônica	39
3	Trabalhos Relacionados	42
4	Técnica de Predição de Redshift Fotométrico	47
4.1	Notação	47
4.2	Motivação	48
4.3	Discussão	49
4.4	Alteração de exemplos com base nos erros	50

4.5	Exemplo numérico	54
5	Experimentos	56
5.1	Configuração de <i>hardware</i> e <i>software</i>	56
5.2	Conjuntos de Dados	57
5.3	Análise Exploratória e Pré Processamento dos Dados	60
5.3.1	SDSS	61
5.3.2	COIN/Teddy	72
5.3.3	COIN/Happy	76
5.3.4	Divisão dos conjuntos de dados	81
5.4	Métricas de Avaliação	81
5.4.1	MAD	81
5.4.2	Taxa de valores extremos (<i>outliers</i>)	82
5.5	Regressores usados nos experimentos	82
5.5.1	Configurações relacionadas às redes neurais artificiais	84
5.6	Resultados	85
6	Conclusões	99
6.1	Análise Retrospectiva	99
6.2	Trabalhos Futuros	101
	Apêndices	104
A	Resultados com outras Métricas	104
A.1	MSE, MAE, RMSE	104
A.1.1	Coeficiente de determinação	105
A.2	Resultados	106
B	Resultados com outras Quantidades de Épocas	110
	Referências	120

1- Introdução

1.1- Contextualização

A Astronomia está entre as primeiras disciplinas científicas a experimentar a necessidade de manipulação e análise de grandes quantidades de dados, resultantes de recentes levantamentos astronômicos. Novas gerações de instrumentos científicos propiciam o grande aumento desses volumes coletados [Ball and Brunner, 2010; Thompson et al., 2014]. Como exemplo, um dos maiores levantamentos astronômicos realizados até hoje é o *Sloan Digital Sky Survey*. Nesse levantamento, a cada noite, o telescópio produz duzentos *gigabytes* de dados, e até hoje quase um milhão de imagens de campo foram adquiridas, nas quais mais de duzentos milhões de galáxias e ainda mais de estrelas foram detectadas [Kremer et al., 2017].

Nesta nova era rica em dados, a Astronomia e a Ciência da Computação podem se beneficiar mutuamente [Kremer et al., 2017]. Em particular, algoritmos de Aprendizagem de Máquina têm dado suporte à análise de informações relativas a eventos astronômicos. Técnicas de aprendizagem de máquina têm sido usadas para melhorar os parâmetros de entrada de um sistema, filtrar dados durante a leitura, suavizar ou eliminar interferências obtidas pelos sensores e telescópios durante a análise nos espectros mensurados [Thompson et al., 2014].

Um caso de uso particular da aplicação de técnicas de Aprendizado de Máquina em problemas da Astronomia diz respeito a produzir estimativas dos *redshifts* (desvio para o vermelho) fotométricos associados a determinadas galáxias. Quanto mais distante um objeto astronômico se encontra, maior o seu *redshift*, devido à expansão do Universo observada pela primeira vez por Hubble [1929]. Essa expansão, que recentemente se descobriu acelerada, faz com que as ondas eletromagnéticas emitidas por esses objetos sejam deslocadas para comprimentos de onda mais longos por conta do efeito Doppler, o que na faixa visível do espectro eletromagnético é associada ao vermelho.

Um conhecimento mais preciso acerca dos desvios para o vermelho associados a objetos cosmológicos pode levar a um melhor entendimento acerca de como as estruturas

no Universo se formaram e evoluem.

A forma atual mais precisa de estimar desvios para o vermelho é por meio da medição espectroscópica dos objetos de interesse. Porém esse método é muito dispendioso de se aplicar para um grande número de galáxias. Por conta disso, métodos de estimação fotométrica de *redshifts* usando abordagens baseadas em aprendizado de máquina têm sido usados e têm trazido resultados relevantes [D’Isanto and Polsterer, 2017; Gschwend et al., 2017; Comins and Kaufmann III, 2010].

Capturas fotométricas multibandas são realizadas atualmente em grandes escalas, por volta de centenas de milhares de galáxias já foram registradas desta forma. Em contrapartida as capturas espectroscópicas representam uma pequena fração disso, algo em torno de um milhão. Para contornar tal problema buscando a completude desse catálogo, o uso de estimadores de *redshifts* fotométricos (foto-z) têm sido uma ferramenta muito útil nesse esforço [Reis et al., 2012].

Métodos baseados em Aprendizado de Máquina para estimação de desvio para o vermelho presumem a existência de um conjunto de dados de treinamento a ser usado no ajuste dos parâmetros do modelo de aprendizado de máquina durante a fase de aprendizado. No caso particular da estimação de desvio para o vermelho, cada entrada nesse conjunto de dados de treinamento corresponde a observações feitas para uma galáxia, quasar ou estrela. Cada entrada é composta por valores de magnitude observados para o objeto, além de respectivas estimativas de erro para cada magnitude. As medidas de magnitude estão relacionadas com os erros de forma que quanto maior o erro da leitura, menos confiável é o valor de magnitude obtido. Isto se deve ao fato de que quanto mais estiver afastado o objeto, mais fraco será o seu brilho. Em cada entrada, também há o valor relativo ao próprio *redshift* espectroscópico. Nesse contexto, o objetivo de usar um algoritmo de aprendizado de máquina é produzir um modelo preditivo, que possa ser usado para estimar o desvio de outros objetos astronômicos no mesmo sistema fotométrico.

Nesta dissertação, propomos um método baseado em Aprendizado de Máquina para gerar modelos de estimação de *redshifts* fotométricos. Esse método é proposto no contexto de uma técnica particular de Aprendizagem de Máquina (AM), as redes neurais artificiais.

A ideia básica do método proposto consiste em utilizar as medidas de erro relacionados às magnitudes para calibrar a confiança do modelo na observação de magnitude

correspondente. Intuitivamente, se uma medida de magnitude está associada a uma estimativa de erro grande, o modelo preditivo deveria dar menos importância a esse valor de magnitude durante a computação da estimativa do desvio.

1.2- Justificativa

Modelos preditivos treinados a partir de um conjunto de dados com a finalidade de determinar uma relação funcional entre *redshifts* e cores são aplicados às imagens de galáxias de interesse para estimar seus valores correspondentes. Esses métodos incluem ajustes polinomiais, vizinhos mais próximos, redes neurais, processos gaussianos, etc [Sánchez et al., 2014; Bilicki et al., 2017; D’Isanto and Polsterer, 2017].

Uma característica particular dos conjuntos de dados astronômicos relacionados a desvios para o vermelho é que cada magnitude medida para um determinado objeto (galáxia ou quasar) é acompanhada de seu erro correspondente, que por sua vez é uma estimativa de quão confiável é a medição de magnitude. Duas abordagens simples para ajustar os modelos de estimativa de *redshift* fotométrico são (1) desconsiderar a existência das medidas de erro, e (2) tratar essas medidas como se estivessem no mesmo espaço conceitual dos valores de magnitude. A primeira abordagem, embora de simples implementação, desconsidera informação potencialmente relevante para o treinamento do modelo de aprendizado de máquina. A segunda abordagem também é limitada na medida em que considera medidas de magnitude e de erro como conceitualmente equivalentes e não relacionadas.

Este trabalho propõe uma nova técnica de treinamento de redes neurais para estimativa de *redshift* fotométrico, que é inspirada na técnica de regularização de modelos de redes neurais artificiais denominada *dropout* [Srivastava et al., 2014]. Essa proposta considera as medições de erro para orientar o processo de treinamento da rede neural artificial de forma que um sinal de magnitude possa ser desativado em virtude de seu erro associado. Quanto maior o erro, maior a probabilidade de um sinal de magnitude ser desativado.

Vamos discutir essa técnica com mais detalhes nos Capítulos 4 e 5.

1.3- Objetivos

O objetivo deste trabalho é propor uma técnica baseada em aprendizado de máquina para estimar *redshifts* fotométricos. A pergunta de pesquisa que este trabalho se propõe a investigar está relacionada ao uso dos erros das magnitudes para guiar o treinamento para medidas de *redshifts* fotométricos. Para cada leitura das magnitudes nas bandas, há um valor de erro associado relacionado com sua confiabilidade. Quanto maior o valor do erro medido, menor é a confiabilidade do sinal magnitude em questão. Nesse contexto, queremos avaliar se as informações sobre os erros das magnitudes podem ser usadas para melhorar o desempenho preditivo de modelos de estimativa de *redshifts* fotométricos. Nossa hipótese é que usar essas informações de erros para guiar o processo de treinamento de um modelo de rede neural artificial pode resultar em melhores modelos preditivos para o problema de estimação de desvios. A intuição que investigamos aqui é que, quando uma observação de magnitude apresenta uma medida de erro correspondente muito distante do erro esperado para aquela magnitude, isso é um indício de que a confiança do modelo preditivo naquela observação deveria diminuir. Os objetivos específicos são descritos a seguir.

- O primeiro objetivo específico é definir e validar uma técnica de treinamento de redes neurais artificiais que utiliza de forma estratégica os erros associados a magnitudes.
- Posto que a técnica supracitada é uma função dos erros observado e esperado para cada banda de magnitude, outro objetivo específico é investigar técnicas de aprendizado de máquina para ajustar uma curva que permita determinar o erro esperado para um determinado valor de magnitude.

1.4- Metodologia

Para os experimentos, utilizamos os dados de leituras fotométricas publicamente disponíveis pelo Programa *Cosmostatistics Initiative* (COIN) e pelo portal do *SciServer*. No programa COIN os dados são construídos a partir da amostra espectroscópica do

telescópio *Sloan Digital Sky Survey* (SDSS)¹. Os conjuntos de dados disponibilizados possuem duas subamostras: (1) *Happy*, projetada para isolar o efeito de cobertura espectroscópica limitada da amostra no espaço cor/magnitude; (2) *Teddy*, projetado para reproduzir o efeito de distintas distribuições de erros fotométricos e sua convolução com cobertura espacial cor/magnitude entre as amostras espectroscópicas e fotométricas. Em todos os conjuntos de dados aqui apresentados (COIN/SDSS), cada objeto de exemplo contém magnitudes nas cinco bandas (ugriz) e seus respectivos erros junto ao valor de *redshift* espectroscópico correspondente. Os valores de *redshift* são os valores alvo, e os demais são as dimensões da solução do modelo a ser treinado, totalizado em dez dimensões, no caso de magnitudes mais erros. Para a construção dos modelos, os dados foram separados em conjunto de treinamento, validação e teste. Em relação aos dados do COIN, cada subamostra (*Teddy* e *Happy*) possuem cada uma quatro conjuntos de dados denominados A, B, C e D fornecidos. O conjunto A tem por finalidade ser o conjunto de treinamento, e os demais (B, C e D) serem os conjuntos de testes.

Para o desenvolvimento dos modelos de aprendizagem de máquina usamos o *Python* como linguagem base. Usamos o *Tensorflow* e a biblioteca *Scikit-Learn* nesse desenvolvimento.

Outro modelo que exploramos além das redes neurais artificiais, é o de regressão gerado pelo algoritmo XGBoost [Chen and Guestrin, 2016], utilizado como base de referência (*baseline*) para guiar os resultados. Para avaliar a efetividade na tarefa de predição de *redshifts* fotométricos, são realizados experimentos computacionais comparativos usando diferentes *datasets* e funções de custo.

Para a avaliação dos modelos, utilizamos as métricas de erro médio quadrado (MSE - *mean squared error*), raiz do erro médio quadrado (RMSE - *root mean squared error*), erro médio absoluto (MAE - *mean absolute error*), coeficiente de determinação (R2 - *coefficient of determination*), desvio absoluto mediano (MAD - *median absolute deviation*) e resíduos com a intenção de se comparar a eficiência entre os modelos.

¹https://github.com/COINtoolbox/photoz_catalogues

1.5- Organização dos Capítulos

O restante desta dissertação está organizado conforme descrição a seguir. O Capítulo 2 apresenta a fundamentação teórica. O Capítulo 3 apresenta os trabalhos relacionados. O Capítulo 4 apresenta nossa proposta de regularização sensível a erros para predição de *redshifts* fotométricos. O Capítulo 5 descreve os experimentos computacionais realizados para validação da nossa proposta de predição. O Capítulo 6 expõe as conclusões, com uma análise retrospectiva acerca dos resultados obtidos e das contribuições alcançadas, assim como apresenta descrições de possíveis continuações da pesquisa. O Apêndice A e o Apêndice B apresentam resultados experimentais complementares aos apresentados no Capítulo 5.

2- Fundamentação Teórica

Este capítulo apresenta a fundamentação teórica necessária para este trabalho. A Seção 2.1 descreve os conceitos relevantes da Astronomia. A Seção 2.2 discorre sobre os conceitos relacionados a métodos de regressão em aprendizagem de máquina considerados nesta dissertação.

2.1- Conceitos Relevantes da Astronomia

Esta seção está organizada da seguinte forma: a Seção 2.1.1 discorre sobre as características do tipo de objeto astronômico de interesse para esta dissertação, a saber, galáxias. A Seção 2.1.2 descreve os conceitos de magnitude e de sistemas de magnitudes. A Seção 2.1.4 refere-se ao efeito *Doppler* e sua relação com *redshift* e a Seção 2.1.5 aborda os conceitos sobre o *redshift* fotométrico e sua importância. Por fim, a Seção 2.1.6 descreve os levantamentos fotométricos e sua importância para a Astronomia atual, com ênfase no SDSS, do qual retiramos os dados utilizados nesta dissertação.

2.1.1- Galáxias

Uma galáxia é um conjunto formado por sistemas estelares, poeira e gás (e matéria escura) ligados gravitacionalmente. De acordo com o sistema de classificação de Hubble, ilustrado na Figura 1, há basicamente 4 tipos de galáxias: espirais e espirais barradas, lenticulares, elípticas e irregulares. As galáxias espirais são caracterizadas por possuírem um bojo e um disco (formado por seu momento angular inicial), com braços espirais com diferentes formatos e razão disco/bojo. Nas espirais barradas, o centro pode apresentar uma barra, que não necessariamente é uma estrutura perene da galáxia. Os braços espirais são regiões com concentração de formação estelar, por isso são mais

brilhantes, pois estrelas jovens, brilhantes e quentes (azuis) se destacam.

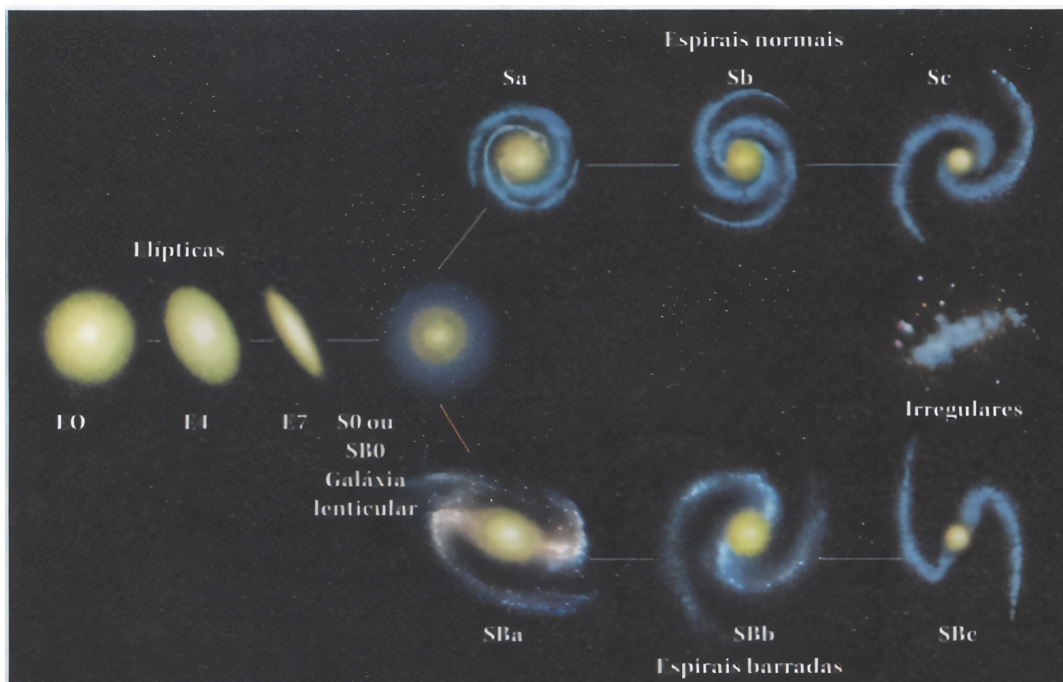


Figura 1 – Esquema de classificação de galáxias elaborado por Hubble. Retirado de Comins and Kaufmann III [2010].

A Via Láctea, por exemplo é uma galáxia do tipo espiral barrada. Outras galáxias menores, conhecidas como galáxias satélite, como as Nuvens de Magalhães, orbitam a nossa galáxia.

As galáxias elípticas se caracterizam pela ausência de gás e poeira, tem formato triaxial, sendo suportadas por dispersão de velocidades (diferente do suporte rotacional de espirais) e tendem a ter uma população estelar mais velha (vermelha) e se concentram em ambientes de alta densidade no universo, como aglomerados de galáxias.

As galáxias lenticulares contêm grandes discos ausentes de braços espirais, tornando assim um tipo intermediário entre uma galáxia elíptica e uma galáxia espiral. Esse tipo de galáxia possui pouca formação estelar por já ter consumido ou perdido grande parte do seu material interestelar. As galáxias lenticulares possuem propriedades similares das galáxias elípticas, como estrelas envelhecidas presentes em sua composição.

As galáxias irregulares recebem este nome por não se encaixarem em nenhum dos formatos descritos acima, são altamente distorcidas e completamente assimétricas em formato; a pequena e a grande nuvem de Magalhães são exemplos de galáxias deste tipo, e podem, inclusive, serem vistas a olho nu. Tipicamente são pequenas e ricas em gás, e portanto apresentam regiões de formação estelar recente.

A Via Láctea faz parte do Grupo Local de galáxias, ao qual ainda também pertencem a galáxia de Andrômeda e outras pequenas galáxias satélites. Andrômeda é a maior representante em tamanho no Grupo Local e deve se fundir com a Via Láctea em alguns milhões de anos, apresentando um *blueshift*.

Ondas eletromagnéticas podem ser geradas pelos mais diversos processos físicos em diferentes frequências (ver legenda na Figura 2). Elétrons acelerados em nuvens de gás quente emitem em rádio. Do núcleo de estrelas, são emitidos raios gama produzidos por reações de fusão nuclear. Esses raios gama são espalhados na atmosfera estelar por diferentes elementos químicos, os quais suas diferentes transições fazem surgir frequências presentes na faixa que vai, essencialmente, do infravermelho ao ultravioleta. Isso é possível por meio de linhas de absorção associadas às diferentes transições desses elementos químicos. A frequência do pico da emissão depende da temperatura das estrelas.

Como já mencionamos, galáxias são um conjunto de estrelas, e sua luz, ou sua Distribuição Espectral de Energia (SED, do inglês *Spectral Energy Density*) vai ser a composição dessas populações estelares, com uma galáxia sendo mais vermelha ou mais azul dependendo se sua população é mais velha ou jovem, respectivamente.

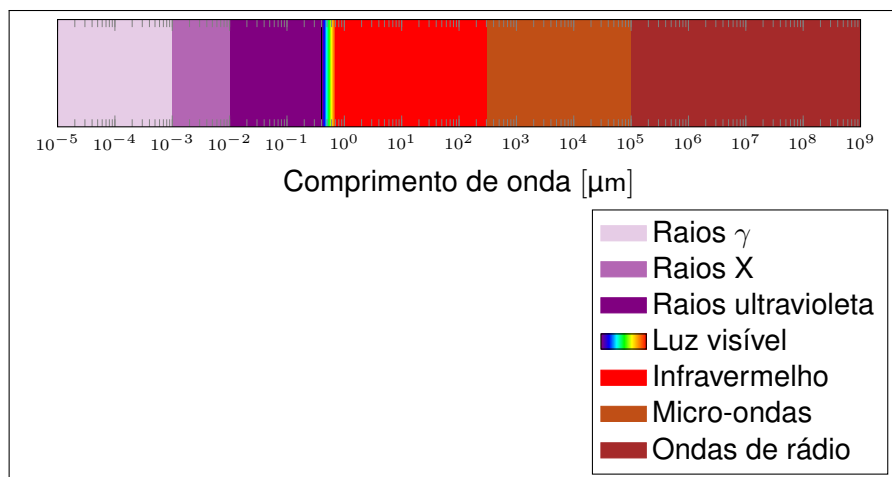


Figura 2 – Espectro eletromagnético. Créditos da figura: Jimmy Westberg¹.

Uma maneira de se explorar essa SED é medir um espectro da galáxia, mas como discutiremos mais a frente, esse método é mais custoso do que simplesmente fazer imagens do céu em vários filtros ou bandas, que são basicamente "vidros coloridos" que deixam passar a radiação em certas faixas de comprimento de onda. O espectro da galáxia se desloca ao longo desses filtros ao sofrer o *redshift* pela expansão do universo.

Isso afeta a medida da SED, medida através da magnitude em algumas bandas, e nos dá uma indicação, grosseira, do *redshift* da galáxia.

2.1.2- Sistemas de Magnitudes

Na Astronomia, desde a antiguidade, define-se o conceito de *magnitude* de um objeto (*e.g.*, estrela ou galáxia), que posteriormente vimos que é proporcional ao logaritmo do fluxo daquele objeto.

Como o brilho (ou magnitude) aparente de um objeto pode variar com a distância, define-se a *magnitude absoluta*, que é a magnitude a uma distância padrão de 10 parsecs (1 parsec = 3,26 anos-luz). Nessa dissertação, a menos que se faça menção em contrário, utilizamos o termo *magnitude* como sinônimo de magnitude aparente.

A escala de valores de magnitude é invertida, no sentido de que, quanto mais brilhante um objeto, menor o seu valor de magnitude. Sendo assim, por exemplo, um corpo celestial de magnitude aparente $m = +1mag$ tem o brilho mais intenso do que uma de magnitude aparente $m = +2mag$. O Sol não foi incluído nessa classificação preliminarmente criada na Grécia antiga. Mas com a evolução da astronomia e redefinição de escalas² foram atribuídos valores negativos de magnitudes para objetos mais brilhantes. O Sol por exemplo tem magnitude de $m = -26,7mag$. Outro exemplo, Sírius, a estrela mais brilhante do céu noturno, hoje tem uma medida de magnitude aparente de $-1,44$.³

Estrelas na Via Láctea com valores de magnitude até um valor próximo a 6 são visíveis a olho nu. Entretanto, telescópios permitem aos astrônomos a observação de galáxias (e de outros corpos celestes) com magnitudes bem maiores (*i.e.*, com brilhos bem menores).

Existem vários sistemas de magnitude usados na Astronomia, com diferentes conjuntos de bandas. O mais famoso por muito tempo foi o sistema Johnson (UBRVI). Com a relevância do levantamento SDSS, o sistema ugriz passou a dominar, sendo usado por vários outros levantamentos subsequentes como CFHTLS, DES, HSC, e LSST.

A Tabela 1 apresenta as diferentes bandas componentes do sistema ugriz junto com os comprimentos de onda médios correspondentes a cada banda.

²com base no fluxo da estrela Vega

³Dados obtidos em <https://earthsky.org>.

Banda	$\bar{\lambda}$
u	365 nm
g	464 nm
r	658 nm
i	806 nm
z	900 nm

Tabela 1 – Sistema de magnitudes ugriz. Para cada letra, é apresentado o comprimento de onda central, denotado por $\bar{\lambda}$.

Além do ugriz há outros sistemas fotométricos que são geralmente caracterizados de acordo com as larguras das bandas passantes. Para cada um dos diversos sistemas fotométricos, um conjunto de objetos estelares padrão primários são fornecidos para a calibração fotométrica e para que sistemas possam ser comparados. Mais detalhes sobre como são calibrados os sistemas fotométricos e como realizar conversões entre eles podem ser encontrados na página do SDSS⁴.

Para ilustrar o cálculo da magnitude em uma banda, a Figura 3 apresenta o espectro de uma estrela quente (linha preta) na perspectiva do sistema de magnitudes ugriz usado pelo Sloan Digital Sky Survey (SDSS). Para cada banda, é calculado o valor do *fluxo* correspondente que passa naquela banda. O fluxo é definido como a quantidade de luz que atinge a área de um detector por unidade de tempo em um telescópio. Essa luz, ao passar pelo filtro, é limitada apenas à região definida pela banda e integrada por sua sensibilidade ao longo da frequência, que não é constante, nem 100%.

Por meio do uso de um sistema de magnitudes específico, astrônomos podem construir uma distribuição espectral de energia grosseira. Para fontes resolvidas e extensas, como galáxias próximas, é possível também que em cada banda a galáxia tenha uma forma distinta, por conta de diferentes populações estelares. Também podemos juntar as imagens em diferentes bandas, como no sistema RGB, para montar imagens coloridas do céu. Como exemplo, considere a Figura 4, que apresenta parte da interface gráfica da ferramenta *SkyServer* disponibilizada no portal do SDSS⁵. Na parte superior direita dessa figura podemos visualizar informações acerca do objetos astronômico selecionado, tais como o tipo, localização no céu (ascensão reta e declinação). Em particular, podemos ler que o objeto selecionado no alto do campo (que é uma imagem composta colorida) é uma galáxia (vermelha) com valores de magnitude $u = 22,60$, g

⁴http://classic.sdss.org/dr4/algorithms/jeg_photometric_eq_dr1.html#usno2SDSS

⁵<http://cas.sdss.org>

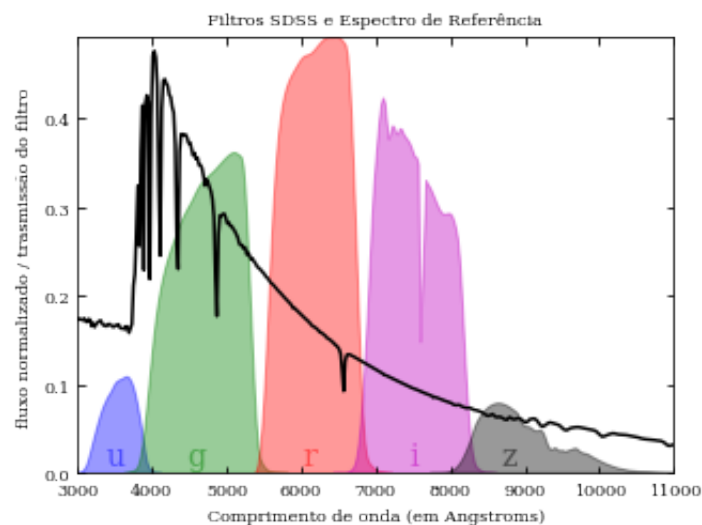


Figura 3 – A curva em preto apresenta o espectro de uma estrela de referência. São também apresentados os diferentes filtros do sistema ugriz. Imagem adaptada de Ivezić et al. [2014].

= 21,80, $r=20,16$, $i = 19,10$ e $z = 18,7$, estando bem perto do limite de magnitudes do SDSS. Essa interface também permite acessar outras propriedades, além do espectro, se houver, do objeto.

Em ciência, uma medida sem seu respectivo erro, tem pouca utilidade. Outra informação disponível para cada objeto contido no catálogo SDSS (embora não seja apresentada na Figura 4) é o erro associado a cada valor de magnitude medido. Como exemplo, a Tabela 2 apresenta os valores de magnitude e de erros para o objeto de identificador 587722981735792679. Como é de se esperar, quanto mais fraco o brilho do objeto em uma banda, maior o seu respectivo erro.

Banda	Magnitude \pm erro
Ultravioleta (u)	$22,60 \pm 0,73$
Verde (g)	$21,80 \pm 0,20$
Vermelho (r)	$20,16 \pm 0,05$
Infravermelho - 7600 Å (i)	$19,10 \pm 0,04$
Infravermelho - 9100 Å (z)	$18,77 \pm 0,08$

Tabela 2 – Para cada objeto no catálogo SDSS, são armazenados os valores de magnitude de acordo com o sistema ugriz, assim como os erros de medição correspondentes.

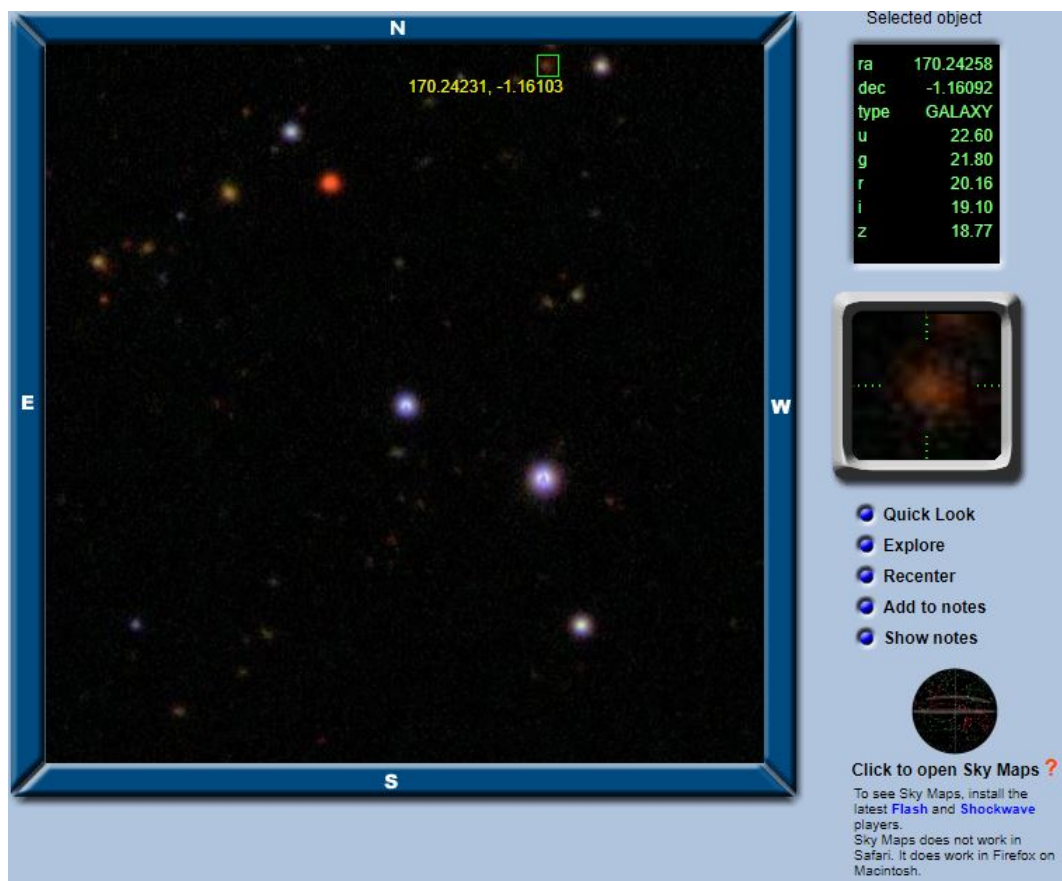


Figura 4 – Tela do SkyServer com informações correspondentes à galáxia cujo identificador é 587722981735792679.

2.1.3- Lei de Hubble

Agora vamos voltar um pouco no tempo. Hubble [1929] apresentou a evidência observacional de que o universo está em expansão. Em particular, Hubble mostrou que um grupo de galáxias próximas à Terra estavam se afastando de nós com uma velocidade proporcional a sua distância, de tal modo que galáxias mais distantes recuam mais rápido do que galáxias próximas [Bahcall, 2015]. A Figura 5 apresenta o famoso Diagrama de Hubble. Esse diagrama apresenta a velocidade observada (no eixo vertical) contra a distância (no eixo horizontal) para galáxias próximas usadas por Hubble em seu artigo de 1929. Medir distâncias é um problema complicado em astronomia, que nesse caso foi resolvido por Henrietta Leavitt ao medir a relação período-luminosidade para estrelas variáveis Cefeidas [Leavitt, 1908], servindo como indicador de distância dessas nebulosas extragalácticas.

Velocity-Distance Relation among Extra-Galactic Nebulae.

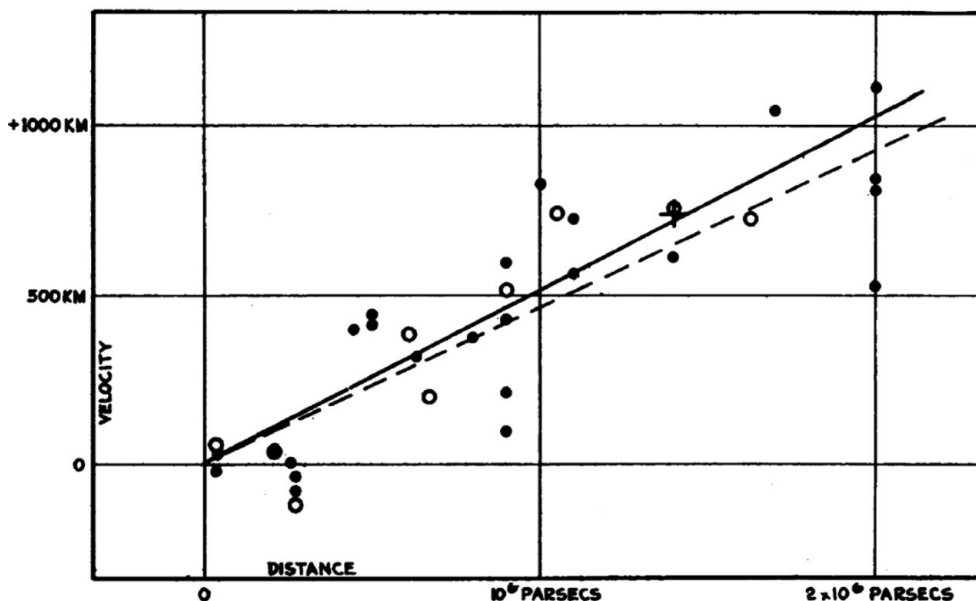


Figura 5 – Diagrama de Hubble, apresentando dados relativos às 24 galáxias (pontos em preto) usadas em sua publicação de 1929. Velocidades estão em Km/s, e distâncias estão em parsecs (1 parsec é igual a 3,26 anos-luz). Imagem retirada de Hubble [1929].

O Diagrama de Hubble revela uma relação linear entre a velocidade de cada galáxia (v) e sua distância (d), conforme a Equação 1, na qual H_0 é conhecido como a constante de Hubble. Como efeito prático dessa relação, se considerarmos duas galáxias, uma a distância d e outra a distância $2d$ da Terra, a segunda galáxia apresenta uma velocidade de recessão duas vezes maior do que a primeira.

$$v = H_0 \times d \quad (1)$$

Na verdade, não há nada de especial em relação ao sistema solar como ponto de referência. De fato, a Lei de Hubble mostra que, em geral, as galáxias estão se afastando umas das outras, culminando na conclusão de que o Universo todo está em constante expansão. Essa conclusão posteriormente levou à teoria do *Big Bang*, de acordo com a qual o Universo começou de um ponto inicial muito pequeno, quente e denso antes de começar a se expandir.

O Universo contém energia e massa desconhecidos que ainda não são simples de detectar por meios de equipamentos e são denominadas de matéria escura (*Dark Matter*), que compõe cerca de 25% da densidade do universo, e a energia escura (*Dark Energy*), que compõe cerca 70%. A energia escura é o nome dado à componente responsável

pela expansão acelerada do Universo [Bachega, 2014].

2.1.4- Efeito Doppler

Efeito *Doppler* é o fenômeno relacionado ao movimento de um objeto emissor de ondas (luminosas ou sonoras) que se afasta ou se aproxima de um observador. Ao se aproximar do observador, há uma compressão da onda, diminuindo o comprimento de onda (e conseqüente aumento da frequência) causando um *blueshift*. Quando o objeto se afasta, o comprimento de onda se alonga, causando o *redshift* [Comins and Kaufmann III, 2010].

Um exemplo do efeito *Doppler* em ação corresponde à situação em que um caminhão de bombeiros está se aproximando de um observador. As ondas sonoras emitidas pela sirene localizada nesse caminhão são percebidas pelo observador em um tom mais agudo do que quando o veículo está em repouso. À medida que o caminhão passa pelo observador e começa a se distanciar, o tom começa a cair e vai ficando mais grave, enquanto o caminhão recua [Gothard and Rosen, 2010].

O efeito *Doppler* se verifica mesmo em tipos de ondas que não precisam de um meio de propagação, como a luz. Quando a fonte está se afastando do observador, a frequência observada é menor do que se a fonte estivesse parada. A diminuição da frequência (ou seja, o aumento o comprimento) da onda faz com que a luz pareça deslocada para a parte vermelha no espectro eletromagnético (ver Figura 2). Uma luz emitida que apresenta *redshift*, é ilustrada conforme na Figura 6. O fenômeno contrário pode acontecer, i.e., quando um objeto está se aproximando de um observador, a luz emitida pelo objeto parece estar deslocada para a parte azul do espectro; nesse caso, diz-se que a luz apresenta *blueshift*.

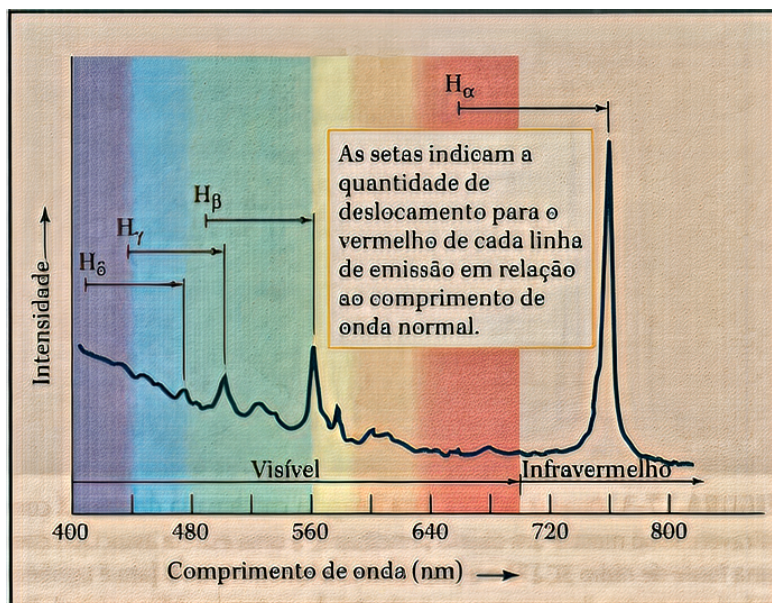


Figura 6 – Se um objeto está se afastando em relação ao observador, o comprimento de onda da luz por ele emitida tende a se alongar, o que a tornará mais avermelhada. Este efeito no contexto da Astronomia é denominado *redshift*. Na figura, a radiação está deslocada para o vermelho cerca de 16% do seu comprimento de onda de repouso. Adaptado de Comins and Kaufmann III [2010]

2.1.5- *Redshift* Fotométrico

Ocorre que o movimento de recessão de um objeto relativamente a uma fonte não é a única causa de desvios para o vermelho. A própria expansão do Universo também faz com que as ondas eletromagnéticas emitidas por objetos distantes (ex.: galáxias e quasares) apresentem esse comportamento, conforme discutimos na próxima seção.

No caso de galáxias distantes, o fato de o Universo estar em expansão se torna o fator preponderante no *redshift* apresentado por aquela galáxia na linha de visada. Movimentos induzidos pela dinâmica da gravidade na vizinhança da galáxia, a chamada velocidade peculiar, também pode contribuir para a velocidade radial medida.

O *redshift* é definido como uma estimativa da velocidade de recessão na direção radial de um objeto astronômico. Existem duas maneiras principais de se medir esse *redshift*, através de espectroscopia e fotometria. Mais especificamente o *redshift* fotométrico é a estimativa baseada no brilho do objeto visto por meio de vários filtros, cada um dos quais detecta a intensidade da luz em uma determinada banda passante. Essa estimativa é importante na Astronomia porque, pela lei de Hubble, é possível determinar

a distância do objeto observado.

Na alternativa, através de espectroscopia, o desvio para o vermelho de uma galáxia distante é medido comparando seu espectro com um espectro de referência obtido em laboratório ou de estrelas ou galáxias próximas com velocidade bem definida. As linhas de emissão e de absorção associadas a transições atômicas ocorrem em comprimentos de onda bem conhecidos. Essas linhas são características de cada elemento químico. Quando as ondas eletromagnéticas provenientes de uma galáxia distante alcançam um telescópio na Terra, as linhas vão estar deslocadas com relação ao comprimento de onda no repouso. Ao medir a localização dessas linhas em espectros, e a diferença com o comprimento de onda (λ) no repouso, o $\Delta\lambda$, os astrônomos podem determinar o desvio para o vermelho das fontes que se afastam, $z = \Delta\lambda/\lambda$, conforme a Equação 2.

$$z = \frac{\lambda_{obs} - \lambda_{rest}}{\lambda_{rest}} \quad (2)$$

2.1.6- Levantamentos Fotométricos

A maneira típica de se explorar o universo é fazer levantamentos astronômicos do maior volume (área e profundidade) possível [Mickaelian, 2016]. Existe um compromisso entre tamanho do telescópio, seu campo de visão - e a sensibilidade do detector - que vai definir o número e o tempo de exposições necessárias para se completar um levantamento. Quanto maior o telescópio e seu campo de visão, mais rapidamente ele pode cobrir o céu, pois precisa de menos apontamentos mais curtos.

O problema da fotometria, em particular feita em um filtro só, é que a informação está projetada em 2D no céu. Para solucionar esse problema e definir a distância de galáxias, por exemplo, podemos usar espectroscopia, já que sabemos, via Lei de Hubble (Seção 2.1.3), que o *redshift* de uma galáxia nos indica sua distância.

Entretanto, surge outra questão. Levantamentos espectroscópicos são mais custosos e demorados do que os fotométricos. Enquanto uma imagem feita de uma exposição de poucos minutos revela centenas de milhares de galáxias, uma exposição espectroscópica da ordem de horas vai observar cerca de milhares de espectros usando

instrumentos multi-fibras (1.000 no BOSS do SDSS [Dawson et al., 2013], e 5.000 para o recém lançado DESI [DESI Collaboration et al., 2016]).

Daí surge a necessidade de se medir distâncias de uma maneira efetiva usando apenas fotometria. Com várias bandas cobrindo uma ampla faixa do espectro, explorando as cores das galáxias, é possível ter uma medida não tão precisa quanto a produzida por espectroscopia, mas ainda assim muito útil para as diversas análises feitas por esse tipo de levantamento.

Alguns dos problemas ao se obter estimativas precisas sobre os dados de foto-z são a falta de cobertura espectroscópica no espaço de cores e magnitudes, e a incompatibilidade nas distribuições de erro entre amostras espectroscópicas e fotométricas, dificultando assim uma análise dos impactos durante as estimativas finais dos valores de foto-z [Beck et al., 2017a]. Amostras de levantamentos tendem a apresentar viés para fontes mais brilhantes, pois é difícil fazer observações espectroscópicas de objetos mais fracos.

O SDSS (*Sloan Digital Sky Survey*)⁶ é um levantamento astronômico que utiliza um telescópio com espelho primário de 2,5 metros de diâmetro acoplado a uma câmera CCD, e a um par de espectrógrafos. Tal dispositivo é capaz de criar mapas tridimensionais muito detalhados do Universo, com imagens multibanda profundas de um terço do céu e no Data Release (DR) 16⁷, para cerca de 200 milhões de galáxias e espectros para quase 3 milhões de galáxias. Sua câmera possui filtros centrados nos comprimentos de onda 354, 476, 628, 769 e 925 nm, chamados respectivamente de u, g, r, i e z. Os dados são liberados em lançamentos anuais (*Data Releases*) por meio do portal *Sciserver*⁸. Nesse trabalho usamos essencialmente versões dos dados do SDSS.

2.2- Métodos de Regressão

Um problema comum na ciência experimental diz respeito a como fazer um conjunto de medições e extrair a melhor estimativa do fenômeno subjacente. No Aprendizado de Máquina, existem diversos métodos cujo propósito é encontrar um conjunto

⁶<https://www.sdss.org/>

⁷<https://www.sdss.org/dr16/scope/>

⁸<https://www.sciserver.org>

de parâmetros que caracterizam um modelo que permita estimar alguma medida de interesse. Em geral, esses métodos utilizam a abordagem dos mínimos quadrados, uma técnica de otimização que procura minimizar a soma dos quadrados dos resíduos das estimativas produzidas por um modelo. Estes métodos são componentes da *tarefa de regressão* em Aprendizado de Máquina. Os modelos resultantes da aplicação desses métodos sobre um conjunto de dados de treinamento são conhecidos como *regressores*.

Algoritmos de regressão são componentes da família de métodos de aprendizado supervisionado, o que significa que esses algoritmos recebem como entrada um conjunto de dados que contém exemplos de treinamento. Cada exemplo de treinamento é composto por um conjunto de medições acerca de um objeto, assim como pelo valor do denominado *atributo alvo*. O objetivo na tarefa de regressão é produzir um modelo que possa prever o valor do atributo alvo sobre medições feitas para novos objetos.

O problema de predição de *redshift* fotométrico pode ser abordado como uma tarefa de regressão em Aprendizado de Máquina. Nessa dissertação, consideramos que cada exemplo de treinamento corresponde observações relativas a uma determinada galáxia. As medições acerca da galáxias são valores de magnitude. O atributo alvo nesse caso é o valor do *redshift* observado para a galáxia.

Esta seção apresenta os métodos de regressão usados nesta dissertação. A Seção 2.2.1 apresenta conceitos sobre as redes neurais artificiais. A Seção 2.2.3 descreve o método que usamos como *baseline* em nossos experimentos, o XGBoost. A própria proposta de predição de *redshift* proposta nesta dissertação (ver Capítulo 4) pode ser parametrizada com algum algoritmo de regressão. A Seção 2.2.2 e a Seção 2.2.4 descrevem os métodos de regressão que usamos para parametrizar a técnica proposta, a saber, algoritmo para indução de árvores de decisão e regressão isotônica.

2.2.1- Redes Neurais MLP

Uma rede neural artificial é um algoritmo de aprendizado de máquina que utiliza unidades computacionais que atuam como um neurônio biológico. Cada unidade computacional, também conhecida como *neurônio artificial*, está conectada a outras unidades, formando uma rede interconectada.

A Figura 7 apresenta a representação de um neurônio artificial. Os n valores de entrada para este neurônio são representados como x_i . Para cada valor de entrada, o neurônio tem um peso associado \bar{w}_i , onde $1 \leq i \leq n$. Outro parâmetro é o viés b . Dentro do neurônio, etapas de pré-ativação e ativação são realizadas. Na pré-ativação, o modelo aplica uma transformação linear, que pode ser definida como uma soma ponderada acrescida de um escalar, conhecido como o viés e denotado por b (Equação 3). A seguir, sua saída z é passada à função de ativação (ou não linearidade) $g(\cdot)$ (Equação 4). A não linearidade permite ao modelo aproximar funções mais complexas, a fim de representar melhor os dados de entrada [Stevens and Antiga, 2019].

$$z = \sum_{i=1}^n \bar{w}_i x_i + b \quad (3)$$

$$\hat{y} = g(z) \quad (4)$$

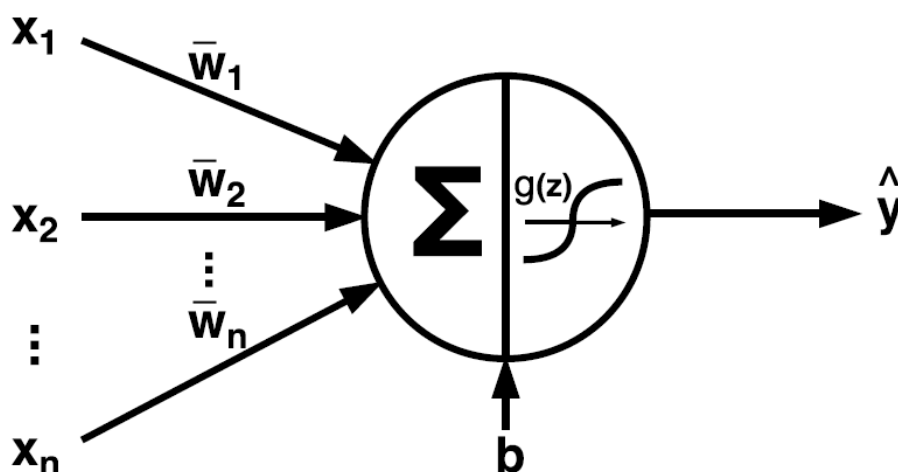


Figura 7 – Neurônio recebe n entradas x_i . São ilustrados os parâmetros ajustáveis (pesos \bar{w}_i e viés b), a função de ativação $g(\cdot)$. Como resultado de sua computação, o neurônio gera um valor \hat{y} . Imagem retirada de Sautermeister [2016].

Um tipo particular de rede neural é conhecido como *Multi Layer Perceptron* (MLP). Em uma rede MLP, neurônios componentes estão organizados em camadas. Cada unidade de uma camada está conectada a todos os neurônios da próxima camada. Valores numéricos conhecidos como *pesos* estão associados a cada conexão entre dois neurônios. Esse pesos correspondem aos parâmetros a serem regulados durante o aprendizado para melhorar o desempenho preditivo do modelo.

A Figura 8 apresenta um esquema de organização típico de uma rede MLP. Em

uma rede MLP, há uma camada de entrada, zeros ou mais camadas ocultas e uma camada de saída. A *camada de entrada* representa os exemplos de treinamento. Cada camada oculta é composta por um conjunto de neurônios que processam os padrões de entrada e produzem padrões de saída. Estes últimos são usados como padrões de entrada para a próxima camada oculta, e assim por diante [Goodfellow et al., 2016]. Por fim, na camada de saída, o resultado final da rede é computado. Cada conexão de um neurônio de uma camada com outro neurônio da camada anterior está associada a um peso. Com uma configuração de pesos apropriada, uma rede MLP com ao menos uma camada oculta pode aproximar qualquer função entre a entrada e a saída [Gardner and Dorling, 1998; Hornik, 1991].

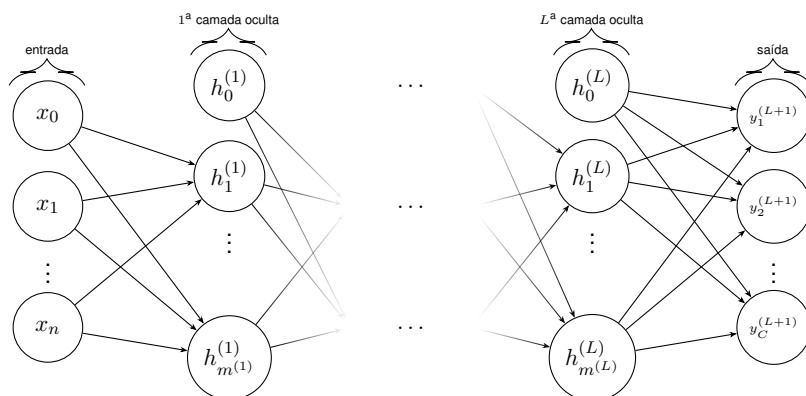


Figura 8 – Esquema de uma rede MLP com $(L + 1)$ camadas. Há n unidades na camada de entrada e C unidades na camada de saída. A l -ésima camada oculta possui $m^{(l)}$ unidades. Imagem retirada de [Bezerra, 2016].

Uma rede MLP gera um modelo preditivo a partir de um conjunto de dados de treinamento. Esta técnica pode ser ajustada a diferentes problemas com a geração de um modelo na etapa de *treinamento*. Nessa etapa, o conjunto de dados de entrada é usado para ajustar os pesos em cada conexão entre neurônios. A diferença entre a saída real e a saída prevista para um determinado exemplo de treinamento é denominada *erro*. Esse erro é utilizado para ajustar os pesos da rede.

A etapa de treinamento de uma rede MLP pode ser dividida em dois passos: passo para frente (*feedforward pass*) e passo para trás (*backforward pass*). Durante o passo para frente, o modelo é alimentado com os dados de entrada para gerar a previsão. O resultado produzido pela rede e o resultado esperado são usados para computar o valor de chamada função de perda (*loss function*). Essa função produz uma medida do erro de previsão produzido pelo modelo. Esse erro é então aplicado para otimizar os

parâmetros (pesos) do modelo. A mudança necessária em cada peso da rede é calculada por meio algoritmo de retropropagação (*backpropagation*) no passo para trás [Stevens and Antiga, 2019]. O processo de aprendizagem durante o treinamento é iterativo, no qual cada iteração é denominada *época de treinamento*. O objetivo desse processo é aprimorar os parâmetros (*i.e.*, pesos) do modelo a fim de minimizar o valor produzido pela função de perda e conseqüentemente produzir um modelo que realize melhores previsões.

A Figura 9 apresenta um exemplo de aplicação de um regressor treinado por rede neural artificial MLP. Esse regressor foi treinado usando 2 camadas ocultas, 26 unidades na primeira camada, 5 unidades na segunda e 200 épocas como hiperparâmetros.

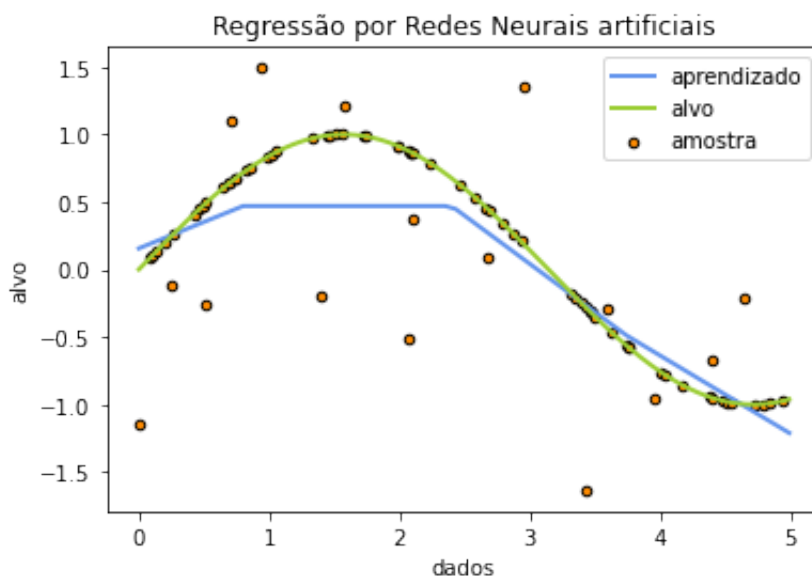


Figura 9 – Exemplo do ajuste de curva do regressor. A linha esverdeada representa o valor alvo, e a linha azul representa o que o modelo aprendeu ou produziu.

2.2.2- Árvores de Decisão

São modelos preditivos de aprendizagem de máquina do tipo supervisionado treinados por meio de algoritmos de árvores. Uma árvore é uma estrutura inspirada em fluxograma onde cada nó representa um teste de resultado verdadeiro ou falso para um valor, cada ramo representa o resultado desse teste, e cada folha representa um valor ou

rótulo de classe. Os possíveis caminhos que levam da raiz à última folha representam regras de regressão ou classificação que um tomador de decisão usará para identificar a estratégia que mais provavelmente alcançará seu objetivo [Rokach and Maimon, 2015].

Os modelos de árvores de decisão são um dos mais rápidos ao se identificar o valor de maior relevância, bem como a relação entre dois ou mais desses valores. Uma árvore de decisão pode ajudar os usuários a criarem novas variáveis, bem como novos recursos para obter mais força na predição do valor de destino. Tais modelos também podem ser usados na fase de exploração de dados. Por exemplo, digamos que o usuário esteja trabalhando em um projeto em que os dados estão disponíveis na forma de várias variáveis. Aqui, a modelo irá prever a variável mais significativa com facilidade [Sullivan, 2017].

As árvores de decisão usadas em uma tarefa de regressão usam uma métrica de impureza adequada para variáveis contínuas em vez de índice de Gini e ganho de informação. Além disso, o valor predito por um nó folha é a média dos valores alvo para os exemplos dentro dessa folha. Uma métrica possível é o erro quadrático médio (MSE). Em particular, para quantificar o nível de impureza $I(t)$ em um nó t da árvore de decisão que está sendo construída, podemos aplicar a expressão denotada na Equação 5. D_t é o subconjunto de treinamento no nó t . m_t é o tamanho de D_t (ou seja, o número de exemplos de treinamento no nó t). $y^{(i)}$ é o valor verdadeiro da característica alvo para o i -ésimo exemplo de treinamento em D_t . \bar{y}_t é a média dos valores previstos para exemplos em D_t .

$$I(t) = \text{MSE}(t) = \frac{1}{m_t} \sum_{i \in D_t} (y^{(i)} - \bar{y}_t)^2$$

$$\bar{y}_t = \frac{1}{m_t} \sum_{i \in D_t} y^{(i)}$$
(5)

Na Figura 10 temos um exemplo com dois regressores treinados usando diferentes profundidades.

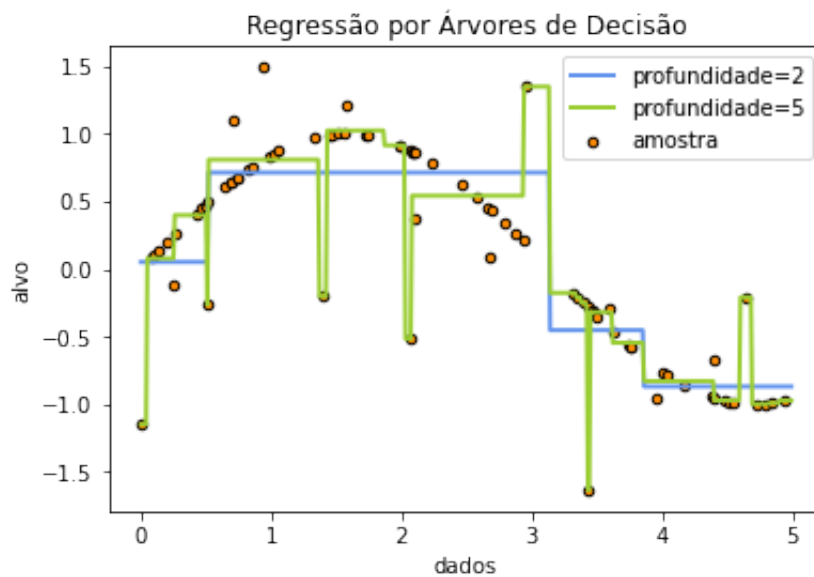


Figura 10 – Exemplo do ajuste de curva de dois regressores por meio de árvores de decisão com profundidades diferentes.

2.2.3- Algoritmo XGBoost

Em aprendizado de máquina, comitês são abordagens para combinar várias técnicas de aprendizado supervisionado para construir um único modelo preditivo. Comitês são aplicáveis às tarefas de classificação e de regressão. Durante o treinamento de um comitê, o objetivo é construir uma coleção de preditores que, juntos, produzem uma predição melhor do que cada um separadamente.

O XGBoost [Chen and Guestrin, 2016] é um algoritmo usado para criar modelos de comitê, formados normalmente por árvores de decisão (ver Seção 2.2.2). O XGBoost é usado para resolver problemas de regressão e de classificação de forma escalonável e flexível, usando técnicas computacionais de paralelismo, padrões de acesso ao cache, compressão e fragmentação de dados para melhorar o tempo de treinamento sem comprometer a acurácia, mesmo para problemas que envolvem muitos dados, contendo por exemplo até bilhões de exemplos [Chen and Guestrin, 2016]. Este sistema é amplamente usado em competições, como por exemplo as que o Kaggle⁹ promove, e como *baseline* para medir a eficiência preditiva de modelos de aprendizagem de máquina.

Em um modelo de comitê típico produzido pelo XGBoost, cada componente é

⁹<https://www.kaggle.com/>

uma árvore de decisão.¹⁰ O XGBoost define uma função objetivo a ser otimizada durante o aprendizado, por meio do método de otimização *Newton-Raphson*, que é similar ao método *Gradient Descent*. Suponha que f_t ($1 \leq t \leq T$) representa uma das T árvores componentes de um modelo comitê produzido pelo XGBoost. Cada árvore é um elemento do conjunto \mathcal{F} , o espaço de árvores compatíveis com o conjunto de treinamento utilizado, $f_t \in \mathcal{F}$. O valor previsto produzido pelo comitê de árvores para o exemplo de treinamento $x^{(i)}$ é dado pela Equação 6.

$$\hat{y}^{(i)} = \sum_{k=1}^T f_k(x^{(i)}) \quad (6)$$

Ajustar todas as árvores do comitê de uma única vez é intratável computacionalmente. Por conta disso, o XGBoost usa um procedimento de otimização aditivo, no qual ocorre a otimização de uma árvore por vez.

$$\begin{aligned} \hat{y}_0^{(i)} &= 0 \\ \hat{y}_1^{(i)} &= f_1(x^{(i)}) = \hat{y}_0^{(i)} + f_1(x^{(i)}) \\ \hat{y}_2^{(i)} &= f_1(x^{(i)}) + f_2(x^{(i)}) = \hat{y}_1^{(i)} + f_2(x^{(i)}) \\ &\dots \\ \hat{y}_t^{(i)} &= \sum_{k=1}^t f_k(x^{(i)}) = \hat{y}_{t-1}^{(i)} + f_t(x^{(i)}) \end{aligned} \quad (7)$$

Durante o aprendizado, o XGBoost procura minimizar uma função objetivo. É possível provar ([Chen and Guestrin, 2016]) que a função objetivo a ser otimizada para selecionar a t -ésima árvore é a apresentada na Equação 8.

$$\text{obj}_t = \sum_{i=1}^m [g_i f_t(x^{(i)}) + \frac{1}{2} h_i f_t^2(x^{(i)})] + \Omega(f_t) \quad (8)$$

Na expressão acima, m é a quantidade de exemplos no conjunto de treinamento. A expressão $\Omega(f_t)$ é um termo de regularização para incentivar o processo de otimização a evitar a produção de modelos que se ajustam demasiadamente aos dados de treinamento. As funções g_i e h_i são as derivadas parciais de primeira e de segunda ordem da função l com relação a $\hat{y}_{i-1}^{(i)}$.

¹⁰A descrição do funcionamento do XGBoost fornecido aqui é baseada em <https://xgboost.readthedocs.io/en/latest/tutorials/model.html>.

$$g_i = \partial_{\hat{y}_{t-1}^{(i)}} l(y^{(i)}, \hat{y}_{t-1}^{(i)})$$

$$h_i = \partial_{\hat{y}_{t-1}^{(i)2}} l(y^{(i)}, \hat{y}_{t-1}^{(i)}) \quad (9)$$

Na Figura 11 apresentamos um exemplo de dois regressores treinados usando o algoritmo XGBoost. Um dos hiperparâmetros do algoritmo XGBoost é a profundidade máxima das árvores que são induzidas durante o treinamento. A figura apresenta resultados para profundidades 2 e 5.

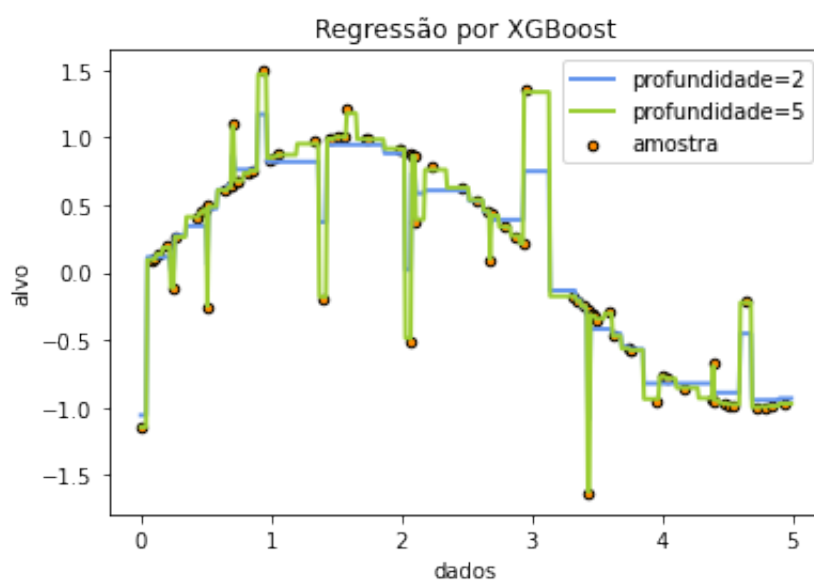


Figura 11 – Exemplo do ajuste para dois regressores XGBoost com profundidades diferentes. Os regressores são representados pelas linhas.

2.2.4- Regressão Isotônica

A regressão isotônica, também conhecida como regressão monotônica, é a técnica de ajuste linear para uma sequência de observações de caráter ascendente para que a curva do modelo passe o mais próximo possível das observações [Barlow, 1972]. A Figura 12 demonstra um exemplo de uma curva ajustada usando um regressor isotônico aplicado a um conjunto de dados fictício gerado por uma função senoidal.

A Figura 13 demonstra outro exemplo de curva ajustada por um regressor isotônico.

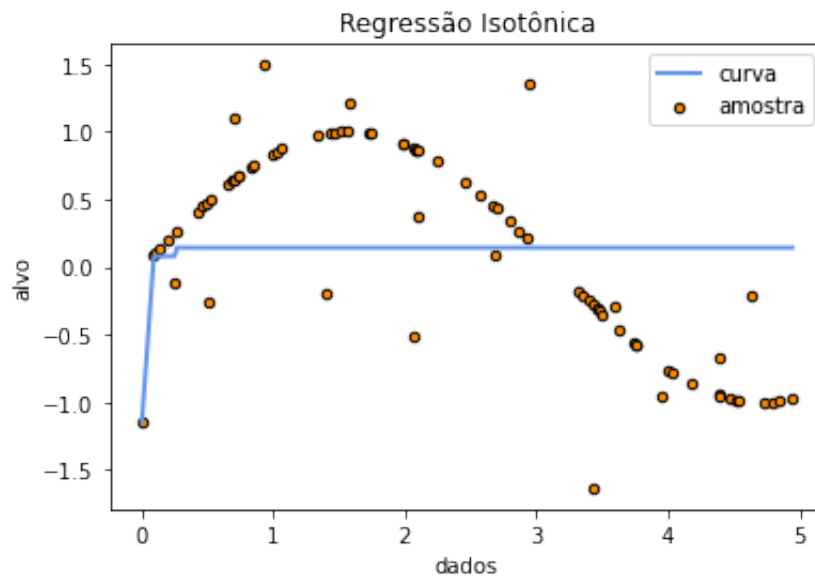


Figura 12 – Exemplo do ajuste de curva por um regressor isotônico. A linha representa a curva ajustada e os pontos são objetos da amostra. Como o conjunto de treinamento apresenta uma tendência não monotônica, o ajuste produzido não é ideal.

O conjunto de dados também fictício gerado por uma função de grau 3.

Na Equação 10 está formalizado o regressor isotônico que envolve encontrar um ajuste de mínimos quadrados ponderados x para um vetor a , com vetor de pesos w sujeito a um conjunto de restrições não contraditórias do tipo $x_i \leq x_j$. E é o conjunto de pares (i, j) para cada restrição $x_i \leq x_j$.

$$\min \sum_{i=1}^n w_i (x_i - a_i)^2 \text{ sujeito a } x_i \leq x_j \forall (i, j) \in E \quad (10)$$

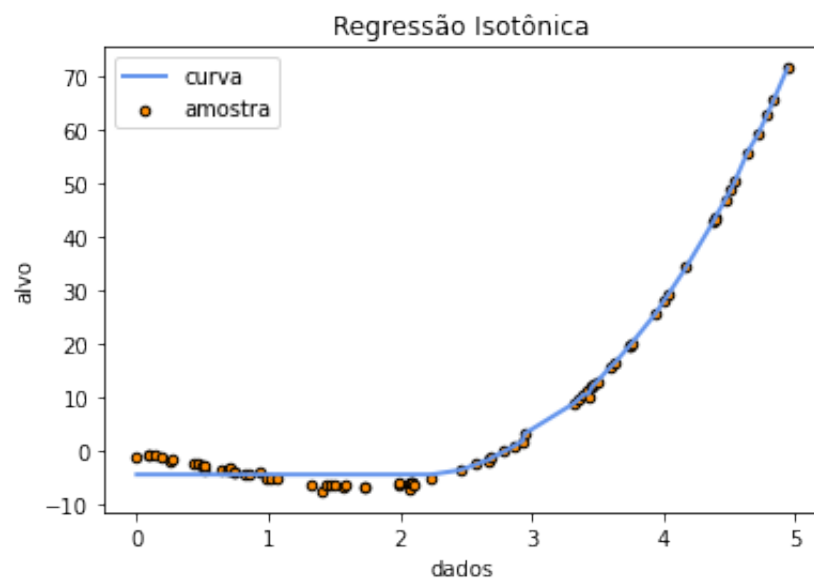


Figura 13 – Exemplo do ajuste de curva por um regressor isotônico. A linha representa a curva ajustada pelo regressor e os pontos são objetos da amostra. Como esta amostragem possui um caráter ascendente, o ajuste está ideal.

3- Trabalhos Relacionados

Para elencar os trabalhos apresentados neste capítulo, usamos a revisão sistemática da literatura como metodologia. Como estratégia de busca utilizamos a ferramenta *Google Scholar*¹ para os seguintes termos principalmente: *redshift*, *galaxies*, *machine learning*, *photometric*. Fizemos uso da técnica tradicional de *snowballing* para alcançar certos trabalhos também. Procuramos incluir os trabalhos mais recentes que obtiveram maiores citações com a finalidade de manter a qualidade das literaturas selecionadas. A maioria dos trabalhos utiliza apenas magnitudes ou respectivas cores no treinamento.

D'Abrusco et al. [2007] usa modelos de redes neurais artificiais para prever foto-z com base nos dados de galáxias provenientes do SDSS-DR4 e SDSS-DR5. O sistema ugriz foi explorado durante os treinamentos dos modelos como características de entrada, porém os erros não foram utilizados.

Sánchez et al. [2014] demonstram que para se calcular os desvios foto-z, as redes neurais artificiais e florestas aleatórias produziram o melhor desempenho nos testes, mas em nenhum dos modelos treinados se utilizou das informações dos erros, apenas as informações sobre magnitudes e cores. Os dados utilizados são provenientes da amostra de Verificação Científica do *DES (Dark Energy Survey)* de 2012-2013.

Hoyle [2016] usa modelos baseados de árvores de decisão (AdaBoost) e redes neurais profundas para prever foto-z com base nos dados de galáxias provenientes do SDSS-DR10. Para os experimentos foram utilizados os valores das magnitudes e das cores, excluindo assim todos dados sobre os erros.

Um método genérico usando redes neurais profundas foi demonstrado em D'Isanto and Polsterer [2017], que permite resolver qualquer tipo de problema de regressão probabilística com base em dados de imagem, como estimar a taxa de formação de estrelas em galáxias. O método proposto é capaz de prever funções de densidade de probabilidade para *redshift* independentemente do tipo de fonte, sendo galáxias, estrelas ou quasares com desempenho preditivo melhor do que os métodos de referência apresentados. Os modelos foram treinados usando os dados de magnitudes e cores

¹<https://scholar.google.com.br>

presentes nos catálogos de quasares e galáxias.

Beck et al. [2017b] usa diversos modelos de aprendizagem de máquina para prever os foto-z. Para tal é usado duas amostragens provenientes do *Sloan Digital Sky Survey (Data Release 12)* (SDSS-DR12) para aferir os impactos de diferentes vieses nas predições, explorando variações espectroscópicas e fotométricas para disparidades em espaço-cor e diferenças nas distribuições dos erros. Nenhum dos modelos treinados se utilizou das informações dos erros, apenas as informações sobre magnitudes e cores.

D'Isanto et al. [2018] propõe um método de seleção de características para avaliar e caracterizar recursos de melhor desempenho para problemas de regressão e classificação. Nesse trabalho são criadas sinteticamente 4520 características por meio da combinação de magnitudes, erros, raios e elipticidades de quasares, retirados dos catálogos SDSS-DR7 e SDSS-DR9, com o objetivo de melhorar o poder preditivo dos modelos de aprendizagem de máquina (Regressão e Classificação). Os erros são usados como entrada para o treinamento dos modelos.

De Wei and Yang [2019] constrói diferentes modelos preditivos como o regressor XGBoost e redes neurais artificiais para estimar os foto-z's. Os dados utilizados são provenientes do SDSS-DR12 e do catálogo *Galaxy Zoo 2*. Todas as variações dos modelos treinados usaram como entrada alternando-se as informações de magnitudes, cores e características construídas a partir destas. O modelo mais complexo possui 14 características como entrada, mas nenhuma delas envolve as informações sobre os erros.

Syarifudin et al. [2019] usa a técnica de aprendizagem de máquina para construir modelos preditivos de foto-z baseado no catálogo do SDSS-DR10. Parte dos modelos treinados usaram metade dos dados do catálogo, e outra parte usaram de forma total com o intuito de avaliar a precisão dos modelos, usando diferentes quantidades de galáxias. Usaram como entrada algumas características das magnitudes para treinar os modelos, mas as informações de erros não foram consideradas.

Neste trabalho usamos técnicas de aprendizagem de máquina para construir modelos preditivos de foto-z baseado nos catálogos do SDSS-DR12 e SDSS-DR16. Usamos os erros não apenas como características, como usados em D'Isanto et al. [2018], mas também como uma abordagem estratégica na etapa de regularização para os modelos de redes neurais.

A Tabela 3 resume as características de cada trabalho pelas funcionalidades apresentadas, como o sistema de magnitude, o uso das informações de erros, faixas de

redshift usados e o tipo de objetos presentes nos catálogos.

A Tabela 4 resume os métodos de aprendizagem de máquina qualificando os algoritmos empregados em cada trabalho. Ao analisar as tabelas podemos conferir que o uso de redes neurais e florestas aleatórias são bastantes populares, e por tanto bastante explorados nos estudos da Astronomia. Em contra partida o uso das informações dos erros durante a execução dos experimentos não são considerados frequentemente.

Autor	Sistema de Magnitude	Erros	Faixa de Redshift	Tipo de Objetos
D'Abrusco et al. [2007]	ugriz	N.A.	0 — 0.5	Galáxias
Sánchez et al. [2014]	ugrizY+cores	N.A.	0 — 1.4	Galáxias
Hoyle [2016]	griz+cores	N.A.	0 — 1.0	Galáxias
D'Isanto and Polsterer [2017]	ugriz+cores	N.A.	0 — 1.0	Galáxias/Quasares
Beck et al. [2017b]	ugriz+cores	N.A.	0 — 1.0	Galáxias
D'Isanto et al. [2018]	ugriz+errors+ radii+ellipticities ²	Feature	0 — 6.0	Quasares
De Wei and Yang [2019]	ugriz+cores ¹	N.A.	0 — 0.8	Galáxias
Syarifudin et al. [2019]	gri	N.A.	0 — 1.0	Galáxias
Este Trabalho	ugriz	Feature+ ³	0 — 1.0	Galáxias

Tabela 3 – Sistema de magnitude demonstra as bandas ou filtros usados como entrada para os algoritmos. A coluna dos erros demonstra como eles foram utilizados nos algoritmos. Os erros marcados como *N.A.* (não aplicado) foram desconsiderados. Os erros marcados como *Feature* significa que foram usados como entradas no conjunto de treinamento. Faixa de *redshift* indica os intervalos do desvio explorado em cada trabalho. A coluna tipo de objetos indica sua natureza, podendo variar entre galáxias e/ou quasares.

¹Além do sistema de magnitude *ugriz+cores*, foram sintetizadas mais por meio destas, chegando até a 14 características para um determinado conjunto de treinamento.

²Além de *ugriz+errors+radii+ellipticities*, foram sintetizadas mais por meio desta, chegando até a 4.520 características para um determinado conjunto de treinamento.

³Além de usarmos os erros como características (*Feature*), também usamos de forma estratégica na regularização, durante o treinamento das redes neurais.

Autor	AdaB ¹	XGB ²	kNN ³	Florestas Aleatórias	Métodos Lineares	Redes Neurais
D'Abrusco et al. [2007]	X	X	X	X	X	✓
Sánchez et al. [2014]	X	X	X	✓	X	✓
Hoyle [2016]	✓	X	X	X	X	✓
D'Isanto and Polsterer [2017]	X	X	X	✓	X	✓
Beck et al. [2017b]	X	X	X	✓	✓	✓
D'Isanto et al. [2018]	X	X	✓	✓	X	X
De Wei and Yang [2019]	X	✓	X	✓	✓	✓
Syarifudin et al. [2019]	X	X	X	X	X	✓
Este Trabalho	X	✓	X	X	✓	✓

Tabela 4 – Trabalhos relacionados por métodos. Cada coluna representa o método de aprendizagem de máquina explorado em cada trabalho.

¹AdaBoost.

²XGBoost.

³K Vizinhos mais próximos.

4- Técnica de Predição de Redshift Fotométrico

Neste capítulo, descrevemos nossa técnica para treinar modelos de redes neurais para predição de *redshift*. Começamos por apresentar a notação e símbolos necessários (Seção 4.1). Em seguida, apresentamos a motivação para a técnica aqui proposta (Seção 4.2). Seguimos então para descrever nossa técnica inspirada em *Dropout* para alterar os exemplos de treinamento com base nos erros de medição disponíveis (Seção 4.4). Em seguida, apresentamos um exemplo numérico de aplicação da técnica proposta (Seção 4.5). Este capítulo termina com uma discussão acerca do conteúdo apresentado (Seção 4.3).

4.1- Notação

Consideramos a disponibilidade de um conjunto de dados para treinamento $\mathcal{D} = \{(\mathbf{x}^{(i)}, z^{(i)})\}$, $1 \leq i \leq |\mathcal{D}|$ de medidas observadas para $|\mathcal{D}|$ objetos astronômicos. Cada objeto é observado por meio de b bandas diferentes.

Cada exemplo de treinamento $\mathbf{x}^{(i)} \in \mathcal{D}$ são pares ordenados cujas componentes seguem o formato esquematizado na Equação 11. Nessa equação, $m_j^{(i)} \in \mathfrak{R}$ denota o valor da magnitude observado na j -ésima banda para o i -ésimo objeto, e $\varepsilon_j^{(i)} \in \mathfrak{R}$ é o seu erro correspondente.

$$\mathbf{x}^{(i)} = \left[m_1^{(i)}, m_2^{(i)}, \dots, m_b^{(i)}, \varepsilon_1^{(i)}, \varepsilon_2^{(i)}, \dots, \varepsilon_b^{(i)} \right] \quad (11)$$

Adicionalmente, para cada exemplo em \mathcal{D} , $z^{(i)} \in \mathfrak{R}^+$ é o valor de *redshift* medido para o i -ésimo objeto. Usamos m_j e ε_j ($1 \leq j \leq b$) para denotar colunas particulares em \mathcal{D} . Repare que $\mathbf{x}^{(i)}$ pode ser visto como a concatenação de dois outros vetores, um de valores de magnitude e outro com valores de error. Usamos $\mathbf{m}^{(i)}$ e $\boldsymbol{\varepsilon}^{(i)}$ ($1 \leq i \leq |\mathcal{D}|$) para denotar os vetores de magnitudes e de erros para o i -ésimo objeto, respectivamente. Em geral, representamos vetores por letras em negrito. Finalmente quando não houver necessidade

de fazer menção específica a alguma banda ou algum objeto em \mathcal{D} , usamos as notações m e ε para denotar valores quaisquer de magnitude e de erro, respectivamente.

4.2- Motivação

Considerando um conjunto de dados de treinamento \mathcal{D} com estrutura conforme descrito acima, existem duas abordagens simples para treinar um modelo de rede neural que produza estimativas para *redshifts* fotométricos. Essas alternativas são descritas a seguir.

1. Usar apenas os valores de magnitudes $m_1^{(i)}, m_2^{(i)}, \dots, m_b^{(i)}$ para treinar o modelo de estimativa, e, portanto, descartar todas as informações relacionadas aos erros observados. Desta forma, a camada de entrada da rede neural seria definida com b unidades.
2. Considerar os valores de magnitude $m_1^{(i)}, m_2^{(i)}, \dots, m_b^{(i)}$ e os erros correspondentes $\varepsilon_1^{(i)}, \varepsilon_2^{(i)}, \dots, \varepsilon_b^{(i)}$. Nesse caso, a camada de entrada da rede neural seria definida com $2 \times b$ unidades.

A desvantagem da primeira abordagem descrita acima é que os erros observados podem transmitir informações úteis para orientar o processo de treinamento. Portanto, não considerá-los durante o treinamento pode resultar em modelos aquém do esperado em termos de qualidade preditiva.

Quanto à segunda abordagem, embora ela considere os erros observados como recursos adicionais para treinar o modelo, o processo de treinamento não é desenhado de forma a explorar explicitamente a relação entre cada valor de magnitude e seu correspondente erro observado. Conforme descrito na Seção 1.3, a hipótese de pesquisa desta dissertação é que os erros associados a valores de magnitudes podem ser usados de forma estratégica durante o treinamento dos modelos de predição.

4.3- Discussão

A abordagem apresentada neste capítulo é inspirada na técnica de regularização denominada *Dropout* [Srivastava et al., 2014]. Essa técnica é frequentemente usada no treinamento de redes neurais artificiais para evitar o sobreajuste (*overfitting*).

A técnica de *Dropout* apresentada em Srivastava et al. [2014] propõe alterar aleatoriamente a arquitetura da rede neural, para minimizar os riscos de que os valores de pesos aprendidos sejam altamente customizados para os dados de treinamento subjacentes e, portanto, não possam ser bem generalizados para dados de teste. Em essência, o *Dropout* simula a combinação de múltiplos modelos sem criar várias redes. Desde sua publicação, essa técnica é amplamente usada durante o treinamento de redes neurais. A Figura 14 ilustra a aplicação dessa técnica sobre unidades de uma rede MLP.

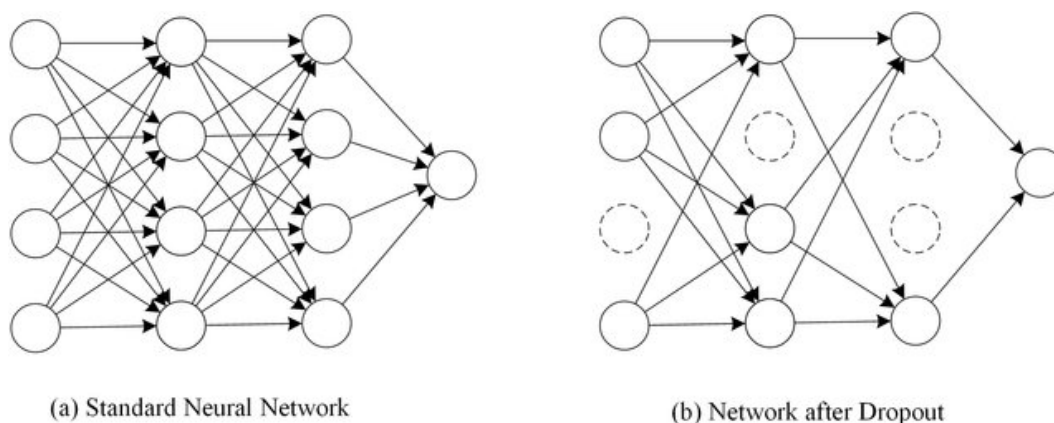


Figura 14 – Visão geral do *Dropout*. Esquerda: rede neural MLP clássica. Direita: durante o treinamento, para cada exemplo de treinamento apresentado à rede, a técnica de *Dropout* seleciona algumas unidades para desligar, no sentido de que essas unidades não participam da computação do resultado gerado pela rede. Imagem retirada de Srivastava et al. [2014].

A técnica de *Dropout* é normalmente aplicada a uma ou mais camadas ocultas durante o treinamento de uma rede neural. Entretanto, o próprio artigo original descreve a possibilidade de aplicar a técnica aos exemplos de treinamento apresentados na camada de entrada da rede, conforme fazemos na técnica proposta neste capítulo.

Durante o treinamento, quando uma camada é configurada com a técnica de *Dropout*, o efeito é que cada neurônio dessa camada é mantido apenas com alguma probabilidade p (que é um hiperparâmetro da técnica). Com probabilidade $1 - p$, a saída

desse neurônio é definida como zero. Mais especificamente, para uma camada com n neurônios, uma máscara de n bits (valores 0 ou 1) é criada. A probabilidade de uma entrada nessa máscara ser igual a 1 é p . Em seguida, essa máscara é aplicada ao neurônios da camada, de forma similar ao que propomos por meio da Equação 15.

Nossa abordagem para alteração dos valores de magnitude de um exemplo de treinamento é inspirada no *Dropout*. Entretanto, nossa abordagem se afasta do *Dropout* clássico no sentido de que não usamos um valor de probabilidade fixo para construir uma máscara de bits para descartar unidades de uma determinada camada. Em vez disso, aplicamos uma máscara de alteração aos padrões de entrada (*i.e.*, aos exemplos de treinamento). Em particular, calculamos a probabilidade de descartar o j -ésimo componente do i -ésimo padrão de entrada a ser apresentado ao processo de treinamento em função de $\varepsilon_j^{(i)}$ e $h_j(m)$.

Outro aspecto relevante é que o procedimento de ajuste de curva para induzir as funções $h_j(\cdot)$ é um problema de regressão univariada, em que as amostras para as variáveis independente e dependente são retiradas das colunas m_j e ε_j , respectivamente. Em princípio, qualquer algoritmo de regressão pode ser usado para produzir o ajuste da curva para $\varepsilon(\cdot, \cdot)$. Definimos esse algoritmo como um hiperparâmetro de nossa técnica.

4.4- Alteração de exemplos com base nos erros

Tomando as limitações descritas acima como motivação, agora descrevemos nossa abordagem para treinar uma rede neural para predição de *redshift*. De forma semelhante à primeira abordagem acima, consideramos apenas as magnitudes m_j como características (*features*) para treinar uma rede neural. No entanto, usamos os erros associados ε_j para decidir quais bandas de magnitude considerar em um determinado exemplo de treinamento.

Dado um valor de magnitude m para uma determinada banda, vamos definir o conceito de *erro esperado* para m . Para auxiliar a descrição desse conceito, considere o gráfico de dispersão apresentado na Figura 15. Cada ponto nesse gráfico representa um par (m, ε) de magnitude e seu erro correspondente em uma determinada banda de medição. Valores de magnitude são representados no eixo horizontal, enquanto valores

de erro são representados no eixo vertical. Note que há uma tendência de aumento do erro de medição, conforme os valores de magnitude aumentam. Estes dados representam exemplos típicos da relação entre magnitudes e erros.

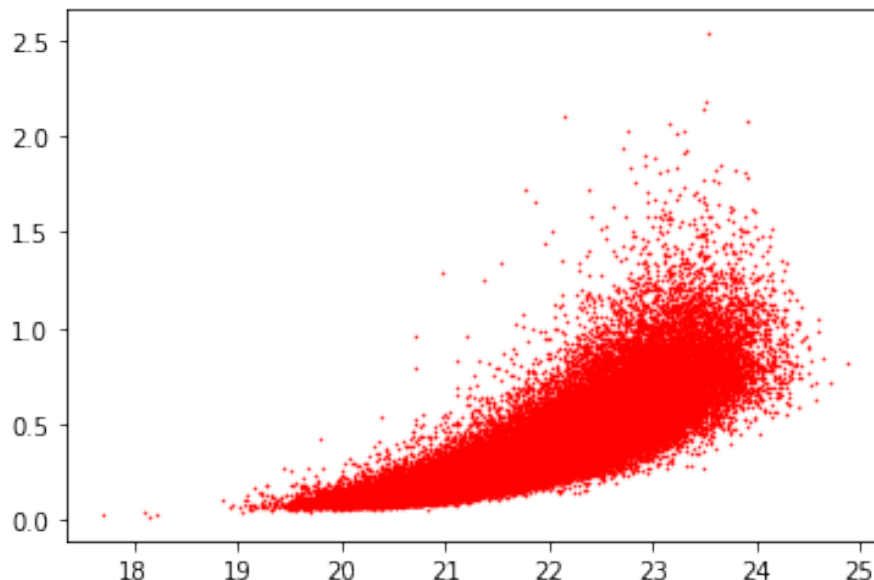


Figura 15 – Os pontos no gráfico correspondem a valores de magnitude (eixo horizontal) e seus respectivos erros (eixo vertical) para uma banda j . Esses pontos vêm da projeção de \mathcal{D} considerando as colunas m_j e ε_j .

Agora considere a Figura 16, a qual apresenta uma curva ajustada aos pontos apresentados por meio da utilização de algum método de regressão. Vamos denotar a função dessa curva por $h_j(\cdot)$. Definimos como erro esperado para o valor de magnitude m como o valor do erro previsto por $h_j(\cdot)$ para o valor m . Usamos o subscrito j para representar o fato de que h_j está associada à coluna m_j .

O cálculo da probabilidade de *Dropout* para a j -ésima banda depende do valor $h_j(m)$, o erro esperado para a magnitude m da banda j . Usamos um procedimento empírico para obter $h_j(m)$. Em particular, $h_j(m)$ é o resultado do ajuste de uma curva ao conjunto de dados bidimensional resultante de uma projeção de \mathcal{D} nas colunas m_j e ε_j . Uma vez que esta curva esteja ajustada, ela pode ser usada para calcular o erro esperado, dada a magnitude de uma banda. A Figura 17 reapresenta o gráfico de dispersão exibido na Figura 16, dessa vez exibindo segmentos de reta verticais conectando a curva ajustada a cada ponto. O comprimento de cada um desses segmentos corresponde ao valor absoluto do numerador na Equação 12.

A intuição que exploramos em nossa abordagem é que, para um determinado

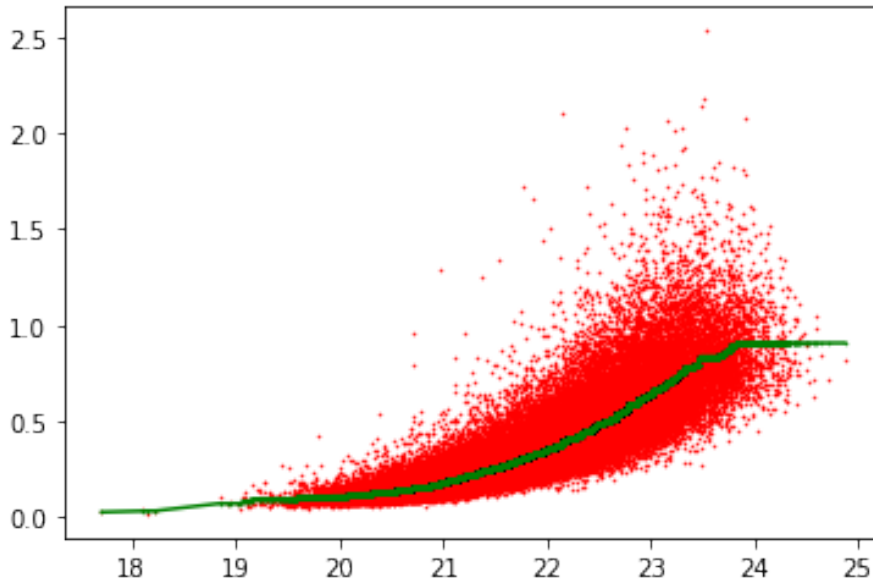


Figura 16 – Curva ajustada aos pontos pontos de magnitude e erro (em verde) corresponde à função $h_j(\cdot)$. Para uma dada magnitude m , $h_j(m)$ corresponde ao erro previsto (ou seja, esperado) para m .

valor de magnitude m , conforme o erro correspondente ε desvia de $h_j(m)$, a confiança no valor m como um preditor do *redshift* deve diminuir. Também consideramos essa intuição no sentido inverso: se a distância entre o erro esperado e o erro medido diminui, a confiança no valor de m como preditor aumenta.

Queremos criar uma expressão para calcular p_{ij} , a probabilidade de descartar $m_j^{(i)}$, o j -ésimo componente do i -ésimo exemplo de treinamento. Como uma etapa intermediária, vamos definir δ_{ij} como na Equação 12.

$$\delta_{ij} = \frac{\varepsilon_j^{(i)} - h_j(m_j^{(i)})}{\varepsilon_j^{(i)}} \quad (12)$$

Denominamos de *desvio* o valor δ_{ij} . Podemos interpretar δ_{ij} como a diferença normalizada entre o erro real medido para o objeto o_i e seu erro previsto (esperado) correspondente na j -ésima banda. Além disso, o valor δ_{ij} está relacionado a p_{ij} , pois quanto maior o valor para δ_{ij} , maior deverá ser a probabilidade do desligamento da banda j para o_i . Para computar p_{ij} a partir de δ_{ij} , usamos a função *softmax*, conforme a Equação 13.

$$p_{ij} = 1 - \frac{e^{\delta_{ij}}}{\sum_{k=1}^b e^{\delta_{ik}}} \quad (13)$$

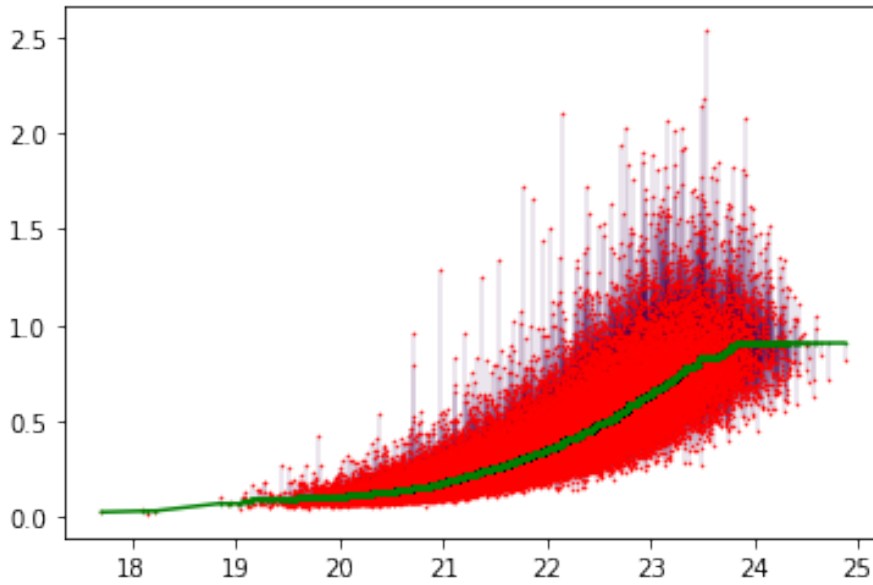


Figura 17 – Cada função $h_j(\cdot)$, $1 \leq j \leq b$ calcula o erro esperado para um determinado valor de magnitude m medido na j -ésima banda. Neste gráfico, os segmentos de reta verticais ligam os pontos de coordenadas $(m, h_j(m))$ e (m, ε) .

Agora descrevemos como produzir $\tilde{\mathbf{m}} = [\tilde{m}_1, \tilde{m}_2, \dots, \tilde{m}_b]$, a versão mascarada de um exemplo de treinamento $\mathbf{m} = [m_1, m_2, \dots, m_b]$ associado a um vetor de erro $\varepsilon = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_b]$. Para isso, suponha que $r_{ij} \sim \text{Bernoulli}(p_{ij})$ indica um valor amostrado de uma distribuição Bernoulli com parâmetro p_{ij} . Sendo assim, $r_{ij} \in \{0, 1\}$. Para calcular $\tilde{m}_j^{(i)}$, $1 \leq j \leq b$, usamos a Equação 14.

$$\tilde{m}_j^{(i)} = m_j^{(i)} \times r_{ij} \quad (14)$$

Se considerarmos que as máscaras computadas para $\mathbf{m}^{(i)}$ estão organizadas em um vetor denotado por $\mathbf{r}^{(i)}$, então podemos reescrever o processo de alteração aplicado a $\mathbf{m}^{(i)}$ conforme a Equação 15. Nessa Equação, \odot denota o produto de Hadamard¹.

$$\tilde{\mathbf{m}}^{(i)} = \mathbf{m}^{(i)} \odot \mathbf{r}^{(i)} \quad (15)$$

¹O produto de Hadamard é uma operação binária que toma dois vetores de mesmas dimensões como operandos e produz outro vetor de mesma dimensão dos operandos, onde cada elemento i é o produto dos elementos i dos dois vetores originais. Essa operação é também definida para matrizes em geral.

4.5- Exemplo numérico

Para esclarecer os passos da proposta apresentada neste capítulo, fornecemos nesta seção um exemplo numérico de sua aplicação. A Figura 18 apresenta uma visão geral desse exemplo para o cálculo da máscara de *dropout* (Seção 4.4) e como ela é aplicada aos exemplos de entrada, seguindo um fluxo ordenado de transformações. Nessa figura, os círculos numerados identificam a sequência de passos para transformação de um exemplo de treinamento em particular.

O passo 1 corresponde a aplicar as funções h_j , $1 \leq j \leq b$ aos respectivos valores de magnitude para gerar os erros esperados. O passo 2 corresponde à aplicação da Equação 12, originando o vetor de desvios $\delta^{(i)}$. Para este passo 2, ocorre previamente o ajuste das funções h_j . No passo 3 é aplicada a Equação 13 para originar o vetor de probabilidades $\mathbf{p}^{(i)}$. Os valores de probabilidade p_{ij} são organizados no vetor $\mathbf{p}^{(i)}$. Durante o passo 4, as componentes do vetor aleatório de máscaras $\mathbf{r}^{(i)}$ são produzidas por meio de amostragem: $r_{ij} \sim \text{Bernoulli}(p_{ij})$. No passo 5, o vetor de saída $\tilde{\mathbf{m}}^{(i)}$ é computado aplicando o vetor de máscaras $\mathbf{r}^{(i)}$ ao vetor $\mathbf{m}^{(i)}$, conforme a Equação 14. Este vetor corresponde ao exemplo a ser usado durante a fase de treinamento do modelo.

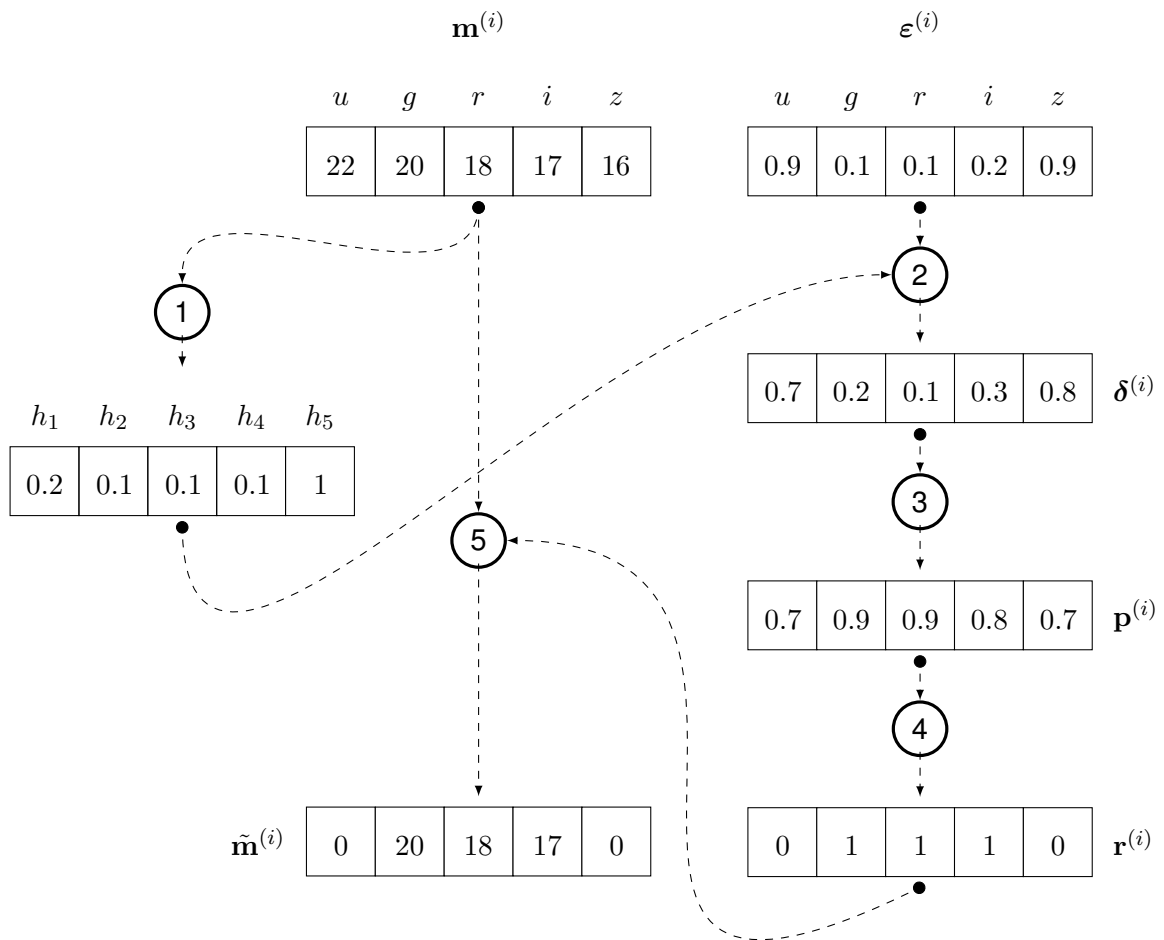


Figura 18 – O passo 1 computa os erros esperados para cada valor de magnitude. O passo 2 aplica a Equação 12 para produzir o vetor de desvios $\boldsymbol{\delta}^{(i)}$. O vetor de probabilidades $\mathbf{p}^{(i)}$ é computado no passo 3 por meio da Equação 13. No passo 4, o vetor aleatório de máscaras ($\mathbf{r}^{(i)}$) é computado. O passo 5 computa o vetor $\tilde{\mathbf{m}}^{(i)}$ aplicando $\mathbf{r}^{(i)}$ ao vetor de magnitudes original, o que resulta no exemplo alterado a ser usado no treinamento.

5- Experimentos

Neste capítulo, descrevemos a construção e execução dos experimentos realizados durante este trabalho. Este capítulo está organizado da seguinte forma. Na Seção 5.1, apresentamos a configuração de *hardware* e *software* utilizados nos experimentos. Na Seção 5.2, detalhamos os conjuntos de dados utilizados, e apresentamos razões para as escolhas dos mesmos. Na Seção 5.3, apresentamos informações resultantes da análise exploratória e pré-processamento desses conjuntos de dados. Na Seção 5.4, descrevemos as métricas usadas para as avaliações de qualidade dos modelos preditivos produzidos. Na Seção 5.5, apresentamos o processo de construção dos modelos e sua utilização na tarefa de prever *redshifts* fotométricos. Na Seção 5.6, apresentamos e analisamos os resultados obtidos com os métodos tradicionais (*baselines*) e com a abordagem que propomos neste trabalho.

5.1- Configuração de *hardware* e *software*

Python é uma das linguagens predominantes na comunidade científica, e na Astronomia não é uma exceção. O Python é uma linguagem multiplataforma gratuita de alto nível, versátil e fácil de programar. Combina programação tradicional, redução de dados e ferramentas de análise de forma amplamente documentada [Faes, 2018]. Toda a implementação foi realizada utilizando a linguagem Python. Para o treinamento dos modelos de redes neurais, utilizamos o *framework* Tensorflow¹ na versão 2.2.0. usando as bibliotecas para o uso de GPU. O TensorFlow é uma plataforma de ferramentas, bibliotecas e recursos abrangente e flexível. Por se tratar de código aberto, permite aos pesquisadores e desenvolvedores criarem e implantarem tecnologias de aprendizagem de máquina de última geração por meio da comunidade. Para treinamento dos modelos envolvendo outros regressores, utilizamos a biblioteca Scikit-Learn² na versão 0.23.1. Em

¹[tensorflow.org/](https://www.tensorflow.org/)

²[scikit-learn.org/](https://www.scikit-learn.org/)

todos os experimentos, utilizamos uma única estação de trabalho, cujo perfil de hardware é apresentado na Tabela 5.

Memória	62.8 GB
GPU	<i>NVIDIA Corp. GP102: GeForce GTX 1080 Ti</i>
CPU	<i>Intel(R) Core(TM) i7-7740X CPU @ 4.30GHz</i>
Núcleos	8
Sistema Operacional	GNU/Linux : 4.15.0-91-generic #92-Ubuntu x86_64

Tabela 5 – Características da estação de trabalho utilizada nos experimentos.

5.2- Conjuntos de Dados

Para os experimentos, utilizamos os dados fotométricos disponibilizados pelo projeto COIN (*Cosmostatistics Initiative*)³, e os disponíveis no portal do *Sciserver*. Todos esses dados são carregados em arquivos no formato CSV, e cada linha desses arquivos representa um objeto que contém informações sobre as cinco bandas de magnitude e seus respectivos erros junto ao valor de *z-spec* correspondente. Os conjuntos de dados do COIN permitem testes que exploram diferentes vieses de amostra usando a fotometria do SDSS [Beck et al., 2017b].

Os dados provenientes do projeto COIN são construídos a partir da amostra espectroscópica do levantamento SDSS-DR12 [Beck et al., 2017b]. Esses dados estão divididos em dois catálogos, denominados *Happy* e *Teddy*. O catálogo *Happy*, projetado para isolar o efeito de cobertura espectroscópica limitada da amostra no espaço cor/magnitude. O catálogo *Teddy*, projetado para reproduzir o efeito de distintas distribuições de erros fotométricos e sua convolução com cobertura espacial cor/magnitude entre as amostras espectroscópicas e fotométricas. Os catálogos acima apresentam os mesmos campos, que estão descritos na Tabela 6. Em cada catálogo, há cinco colunas com valores de magnitude, outras cinco colunas com os erros respectivos de cada valor de magnitude. Os valores de magnitude seguem o sistema de magnitudes *ugriz* [Fukugita et al., 1996] (Seção 2.1.2). Há também duas outras colunas, com o desvio para o vermelho espectroscópico e o seu erro correspondente.

³https://github.com/COINtoolbox/photoz_catalogues

Campo	Tipo	Descrição
ID	<i>int</i>	Identificador do registro
u	<i>float</i>	5 magnitudes do ultravioleta através de g e r para perto de i, e ainda mais perto de z
g	<i>float</i>	Verde
r	<i>float</i>	Vermelho
i	<i>float</i>	Infravermelho
z	<i>float</i>	Z-Infravermelho
uErr	<i>float</i>	Erro de u
gErr	<i>float</i>	Erro de g
rErr	<i>float</i>	Erro de r
iErr	<i>float</i>	Erro de i
zErr	<i>float</i>	Erro de z
redshift	<i>float</i>	Desvio para o vermelho
redshiftErr	<i>float</i>	Erro estimado de redshift

Tabela 6 – Descrição dos campos contidos nos catálogos do projeto COIN.

```

SELECT
  p.objid ,
  p.u,p.g,p.r,p.i,p.z ,
  p.err_u ,p.err_g ,p.err_r ,p.err_i ,p.err_z ,
  s.specobjid ,s.class ,s.z as redshift
FROM
  PhotoObj AS p JOIN SpecObj AS s ON s.bestobjid = p.objid
WHERE
  p.r BETWEEN 0 AND 23 AND s.class = 'GALAXY'
ORDER BY redshift

```

Listagem 1 – Consulta SQL executada na base do *SkyServer*

Além dos dados do COIN, também utilizamos os dados do levantamento *Sloan Digital Sky Survey (Data Release 16)* (SDSS-DR16) presente no portal *Sciserver*. Estes dados, que podem ser obtidos por meio de consultas SQL à base de dados disponível. Estas consultas se tornam possíveis após um breve cadastro no próprio portal do *Sciserver* e acessando a ferramenta *SkyServer Search*⁴. Há duas tabelas de interesse na base de dados, a primeira *PhotoObj* que possui os dados sobre as leituras fotométricas, e a segunda *SpecObj* que possui as leituras dos dados espectroscópicos. Essas tabelas são relacionadas por meio de um identificador numérico que representa um objeto cósmico catalogado pelo levantamento. Submetemos a consulta SQL apresentada na Listagem 1 para obtenção dos dados.

Após esta consulta, o *download* do arquivo contendo o resultado é disponibilizado.

⁴<http://skyserver.sdss.org/dr16/en/tools/search/sql.aspx>

```

SELECT
    p.objid ,
    p.u,p.g,p.r,p.i,p.z,
    p.err_u,p.err_g,p.err_r,p.err_i,p.err_z ,
    s.specobjid,s.class,s.z as redshift
FROM
    PhotoObj AS p JOIN SpecObj AS s ON s.bestobjid = p.objid
WHERE
    p.r BETWEEN 0 AND 23 AND s.class = 'GALAXY'
    AND s.z > $redshift
ORDER BY redshift

```

Listagem 2 – Consulta SQL executada na base do *SkyServer* filtrando por valores de *redshift*.

Porém nem todos os dados estarão presentes. Isso se deve ao fato de que o serviço possui uma limitação técnica de número máximo de registros retornados por consulta. Para contornar essa limitação, utilizamos uma forma de paginar os dados obtendo os registros de forma gradual. Realizamos a paginação ordenando a consulta por meio da cláusula *ORDER BY* no campo *redshift* (tabela *SpecObj* campo *z*). A partir disso, fizemos consultas subsequentes definindo o parâmetro *\$redshift* com o último valor de *z-spec* presente na consulta anterior, conforme a consulta SQL apresentada na Listagem 2.

Executamos a consulta apresentada na Listagem 2 por seis vezes para obtenção de todos os dados possíveis. Após as execuções das consultas, consolidamos os dados em um único arquivo com um total de 2.771.159 registros resultantes da consulta, e suas características são descritas na Tabela 7.

Para este conjunto de dados, filtramos as entradas da banda *r* para excluir os registros maiores do que 23 por ser a magnitude limite do SDSS. Valores de magnitudes maiores que 23 nesta banda não são confiáveis. Outro filtro adotado foi de considerar apenas os objetos classificados como galáxias por só se medir foto-*z* para galáxias e quasares. Estrelas possuem sempre valor de foto-*z* igual a zero por estarem próximas, dentro da nossa própria galáxia. Encontramos valores de magnitude de -9999 para as bandas *u*, *g*, *i* e *z* em alguns objetos da amostra. Isto caracteriza dados espúrios e então removemos esses objetos aplicando um filtro em todos os valores menores que 0 nessas bandas. O total de dados removidos foram de 24 objetos, e em todos eles possuíam -9999 de magnitude ou de erro em alguma das bandas. Posto que os catálogos Happy e Teddy só possuem galáxias e aproximadamente 75.000 exemplos em

Campo	Tipo	Descrição
objid	<i>int</i>	Identificador do registro
specobjid	<i>int</i>	Identificador único de registro como chave estrangeira
class	<i>string</i>	Classificação espectroscópica do objeto espacial Valores possíveis GALAXY, QSO e STAR
u	<i>float</i>	5 magnitudes do ultravioleta através de g e r para perto de i, e ainda mais perto de z
g	<i>float</i>	Verde
r	<i>float</i>	Vermelho
i	<i>float</i>	Infravermelho
z	<i>float</i>	Z-Infravermelho
err_u	<i>float</i>	Erro de u
err_g	<i>float</i>	Erro de g
err_r	<i>float</i>	Erro de r
err_i	<i>float</i>	Erro de i
err_z	<i>float</i>	Erro de z
redshift	<i>float</i>	Desvio para o vermelho

Tabela 7 – Descrição dos valores contidos no conjunto de dados do SDSS-DR16.

cada conjunto A (treinamento), removemos os quasares da amostragem e produzimos uma subamostragem aleatória sem reposição de 120.000 exemplos, com a finalidade de obtermos aproximadamente 75.000 exemplos para o conjunto de treinamento, mantendo assim todos os conjuntos de treinamento explorados com aproximadamente o mesmo tamanho. Os filtros de corte da banda e da classe são observados nas consultas da Listagem 1 e da Listagem 2. A Tabela 8 resume as características dos catálogos que usamos nas análises e experimentos desta dissertação.

Arquivo	Amostras / Exemplos	Tamanho (MB)
SDSS DR16	2.771.159	409,0
COIN/Happy	74.950	9,6
COIN/Teddy	74.309	15,0

Tabela 8 – Características dos arquivos físicos dos conjuntos de dados do SDSS e COIN de forma comparativa.

5.3- Análise Exploratória e Pré Processamento dos Dados

Nesta seção, apresentamos características encontradas durante a análise exploratória realizada sobre os catálogos SDSS-DR16, COIN/Teddy e COIN/Happy. Para cada

conjunto de amostras descrevemos os dados, apresentamos a matriz de correlação e checagem de valores atípicos (*outliers*) com o objetivo de se verificar a integridade das amostras. Esta seção se propõe a responder essas perguntas: Será que as amostras possuem dados consistentes? Será que há muitos dados faltantes ou com ruídos? Quais as dimensões que influenciam mais no valor de foto-z?

5.3.1- SDSS

A Figura 19 apresenta a matriz de correlação produzida a partir do conjunto do SDSS-DR16 logo após a execução das consultas (Listagens 1 e 2), e notamos que para esta amostragem a matriz é mais esparsa comparada aos outros conjuntos (Teddy na Figura 29 e Happy na Figura 33). As dimensões de maiores influências no valor alvo `redshift` são `r` e `err_r`, tornando as demais componentes praticamente irrelevantes. Observamos também forte correlação entre as magnitudes e seus erros correspondentes em cada banda.

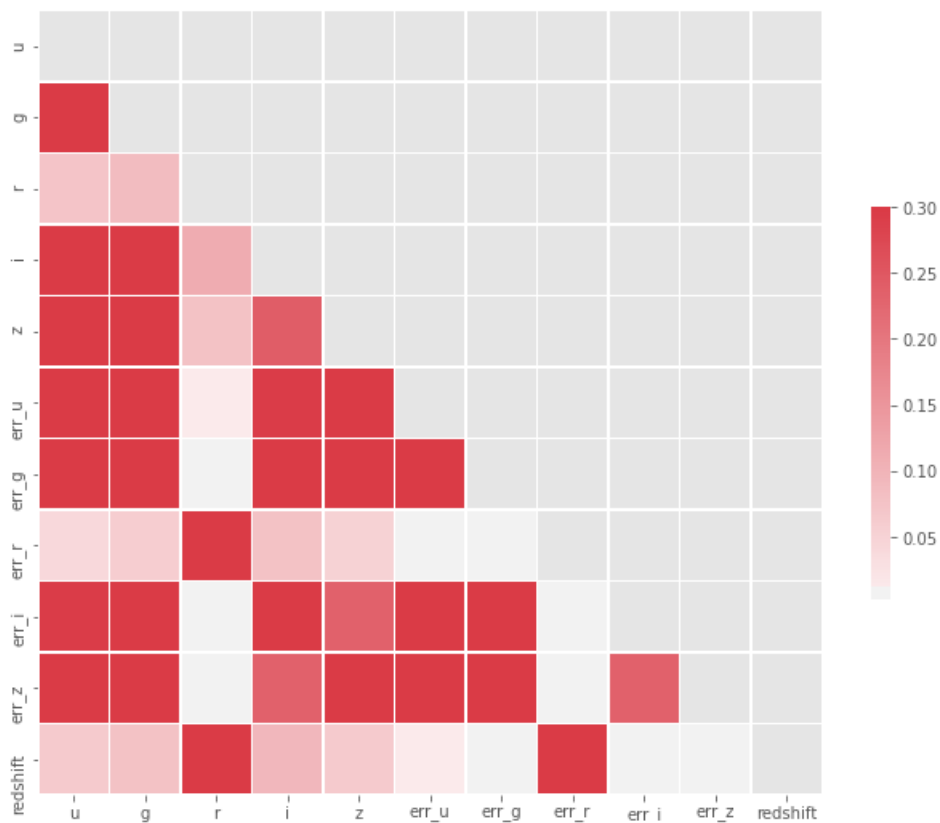


Figura 19 – Matriz de correlação do conjunto SDSS-DR16.

Removemos as colunas `objid`, `specobjid` e `class` presentes nesse conjunto por se tratarem de valores de metadados irrelevantes para os experimentos. Na Figura 20, apresentamos o gráfico de caixas (*box plot*) para o conjunto do SDSS-DR16. Observamos a presença de valores atípicos (*outliers*) nas bandas `u`, `g`, `i`, `z`, e também nos erros `err_u`, `err_g`, `err_i` e `err_z` de -9999, totalizando 24 entradas. Removemos esses objetos aplicando um filtro em todos os valores menores que 0 nas bandas (magnitudes e erros), mantendo um total de 2.771.135 objetos.

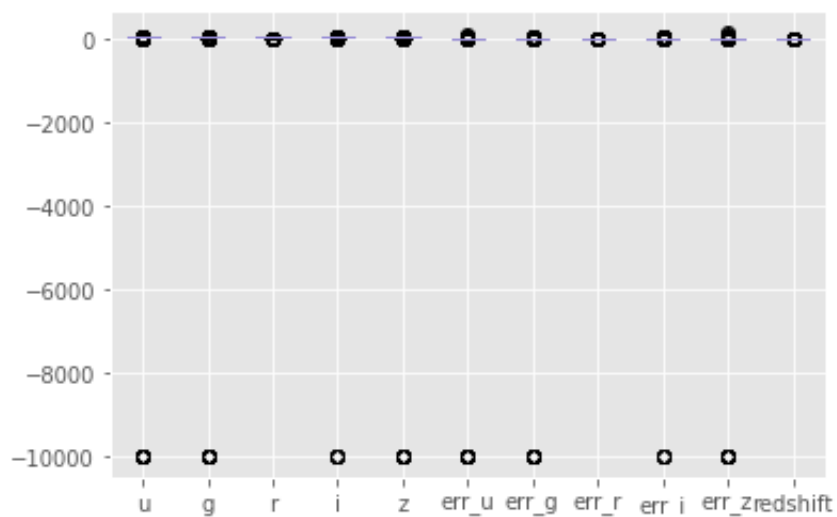


Figura 20 – Boxplot com os valores atípicos do conjunto do SDSS-DR16.

Após a aplicação do filtro, podemos constatar o quanto esses dados influenciaram a amostra, pela Figura 21. Notamos que a matriz de correlação mudou drasticamente aumentando muito a relevância de todas as componentes entre si. Isto comprova que ignorar qualquer desses valores não seria uma decisão adequada na direção de se obter um bom modelo preditivo.

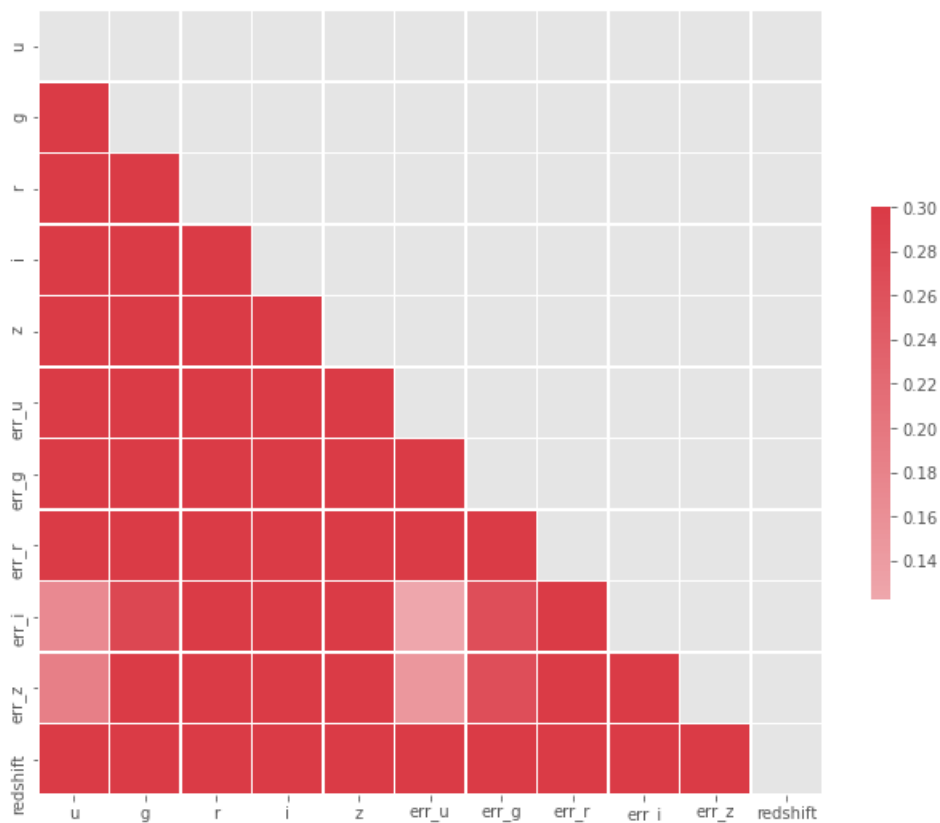


Figura 21 – Matriz de correlação da do conjunto SDSS-DR16 após a remoção dos dados espúrios.

Na Figura 22, apresentamos o gráfico de caixas *box plot* para o conjunto SDSS-DR16 após a filtragem dos dados espúrios. Observamos a presença de valores atípicos (*outliers*) com maior concentração nas bandas i e z, o que pode ser explicado dado que as leituras destas bandas são provenientes de sinais menos brilhantes e, portanto, mais erráticos.

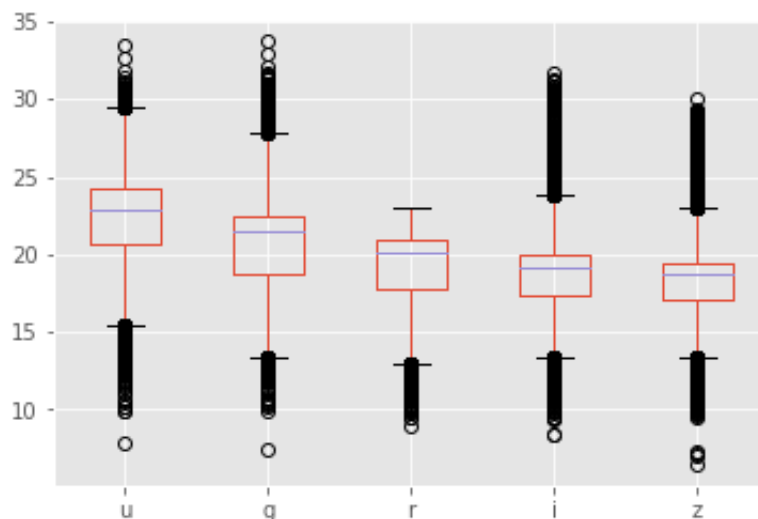


Figura 22 – Boxplot das magnitudes com os valores atípicos do conjunto do SDSS-DR16 após a remoção dos dados espúrios.

Na Figura 23, podemos ver os valores atípicos para as componentes de erros referentes a `err_u`, `err_i` e `err_z` as quais possuem as maiores concentrações de valores atípicos. Este conjunto apresenta a maior amostragem de valores atípicos em relação a Happy (Figura 35) e Teddy (Figura 31).

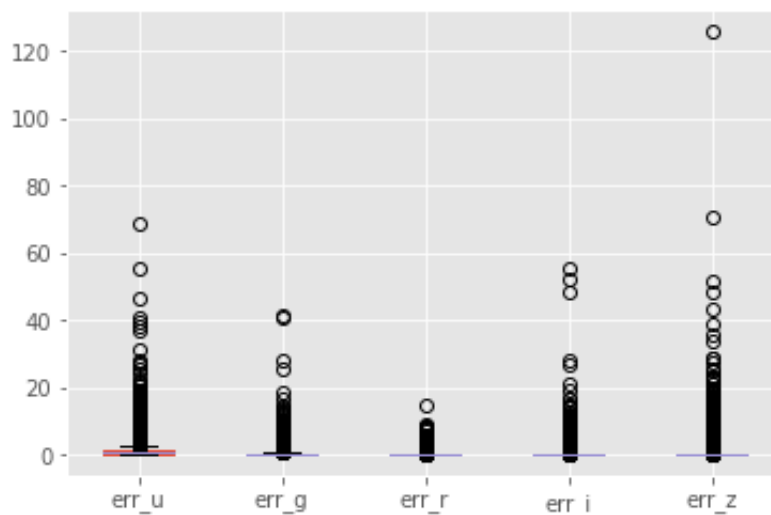


Figura 23 – Boxplot dos erros com os valores atípicos do conjunto do SDSS-DR16 após a remoção dos dados espúrios.

Fizemos a filtragem nos valores de todas as componentes dos erros para todo sinal maior que 1.0, o que resultou na remoção de 756.030 objetos, mantendo assim 2.015.105 objetos na amostra. A Figura 24 demonstra o gráfico de caixas resultante, e na Figura 25 podemos observar o corte com mais expressividade reparando a banda `u`.

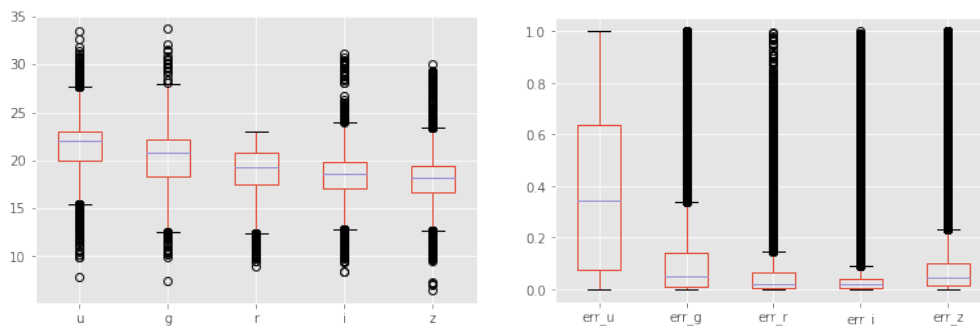


Figura 24 – Boxplot contendo os valores atípicos das magnitudes e dos erros do conjunto do SDSS-DR16 após a filtragem dos valores nas componentes de erros.

Na Figura 25, apresentamos a distribuição das magnitudes pelos erros em cada uma das bandas (u, g, r, i, z) do conjunto SDSS-DR16 após as filtrações realizadas. Nessa figura, podemos observar que há, em geral, uma crescente dos erros em relação à crescente da magnitude, quanto menor o brilho, maior o erro em cada uma das bandas observadas. O sinal tende a ficar mais errático para os objetos menos intensos (brilhantes) em cada banda, e isto é uma característica normal para este tipo de dado astronômico por conta das próprias limitações do SDSS. Ao observar a distribuição da banda u é notada uma característica particular, uma tendência decrescente a partir da magnitude 25 que pode ser explicada por se tratarem de dados espúrios, proveniente de algum provável ruído nas fases de captura dessas amostras.

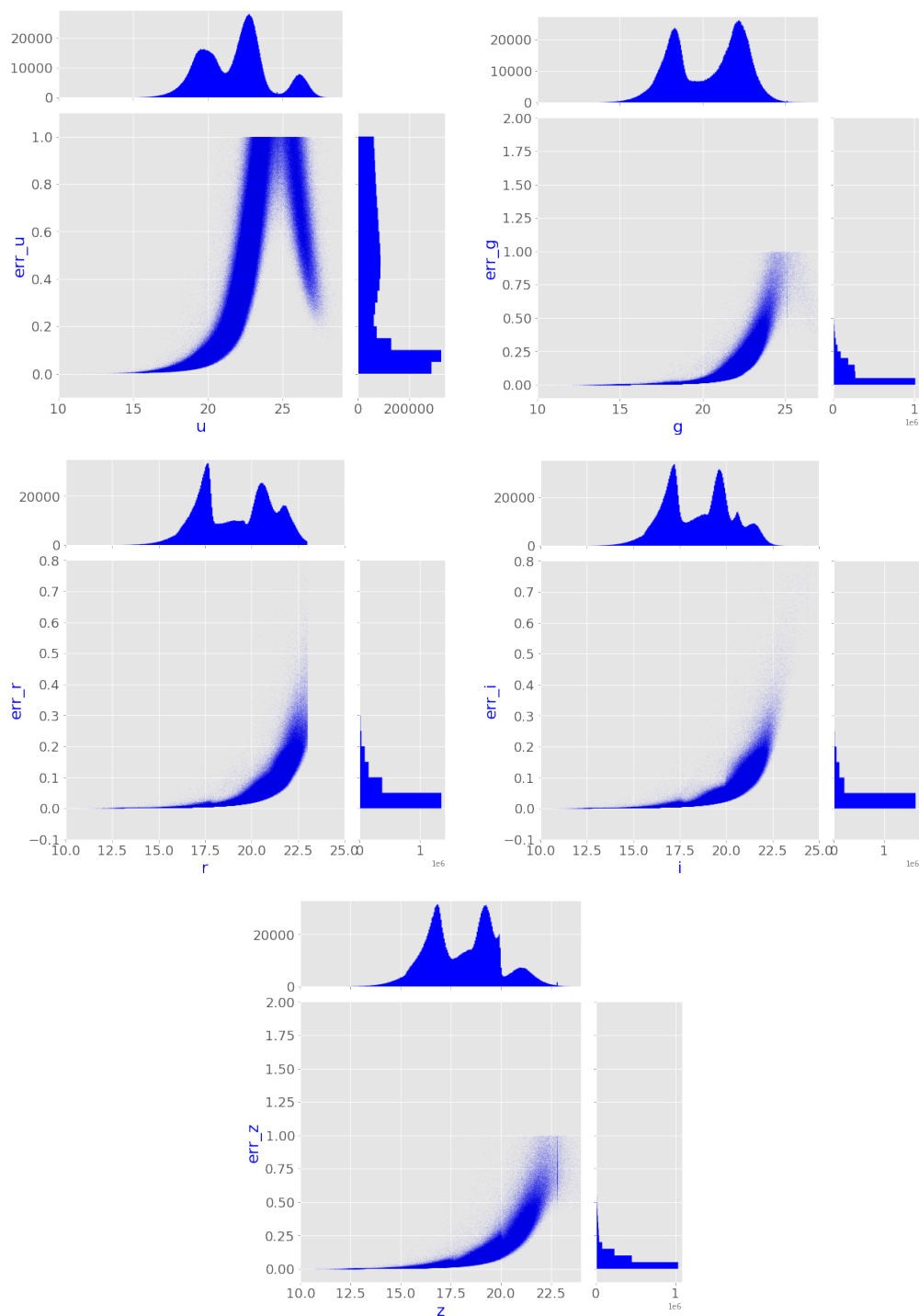


Figura 25 – Distribuição das magnitudes pelos erros em cada banda do conjunto SDSS-DR16 após as filtrações de dados espúrios.

Aplicamos um filtro na banda u para remover os valores superiores a 25, o que resultou em 214.927 objetos removidos, e um total de 1.800.178 na amostra. A Figura 26 apresenta o gráfico boxplot após a filtração.

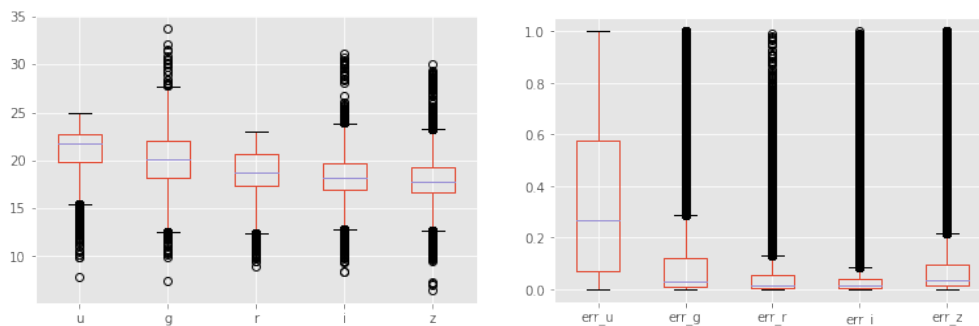


Figura 26 – Boxplot contendo os valores atípicos dos erros do conjunto do SDSS-DR16 após a remoção dos dados espúrios e filtragem na banda u no valor 25.

A Figura 27 apresenta a distribuição das magnitudes pelos erros em cada uma das bandas após a filtragem. Notamos que nenhuma das demais bandas além de u foram afetadas pelo filtro ao se comparar com seu estado anterior (Figura 25).

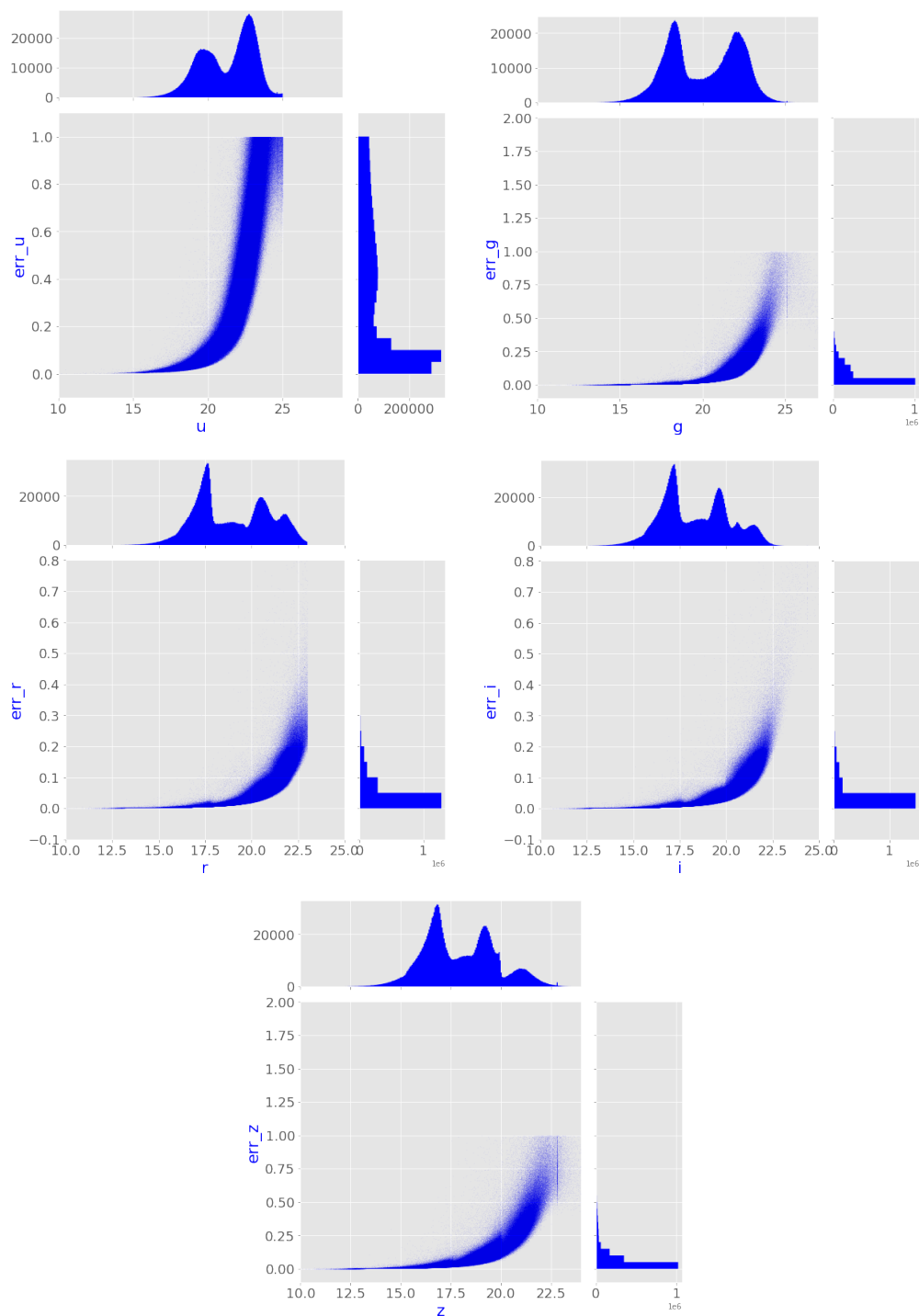


Figura 27 – Distribuição das magnitudes pelos erros em cada banda do conjunto SDSS-DR16 após as filtrações de dados espúrios e filtragem na banda u no valor 25.

A partir desses tratamentos, produzimos uma subamostra (selecionada de forma aleatória e sem reposição) de 120.000 exemplos, com a finalidade de obtermos aproximadamente 75.000 exemplos para o conjunto de treinamento, mantendo assim os resultados

mais comparáveis entre Teddy, Happy e esta agora denominada SDSS-R16-120K.

Na Figura 28, apresentamos a distribuição das magnitudes pelos erros em cada uma das bandas (u , g , r , i , z) do conjunto SDSS-R16-120K. Aqui, podemos observar que geramos uma subamostragem representativa do conjunto geral, se compararmos com a amostra original (Figura 25), vemos que as proporções são mantidas dadas a densidade e a disposição dos pontos, desconsiderando os dados extremos já filtrados.

Quanto aos histogramas apresentados na Figura 28 observamos que há uma descontinuidade no catálogo das bandas u , g por volta da magnitude 21, e nas bandas r , i e z por volta da magnitude 17.5. Este conjunto apresenta a maior descontinuidade em relação ao de Teddy (Figura 32) e Happy (Figura 36). Observamos também que a maior concentração dos erros ocorrem por volta de 0.1 dos erros em cada uma das bandas.

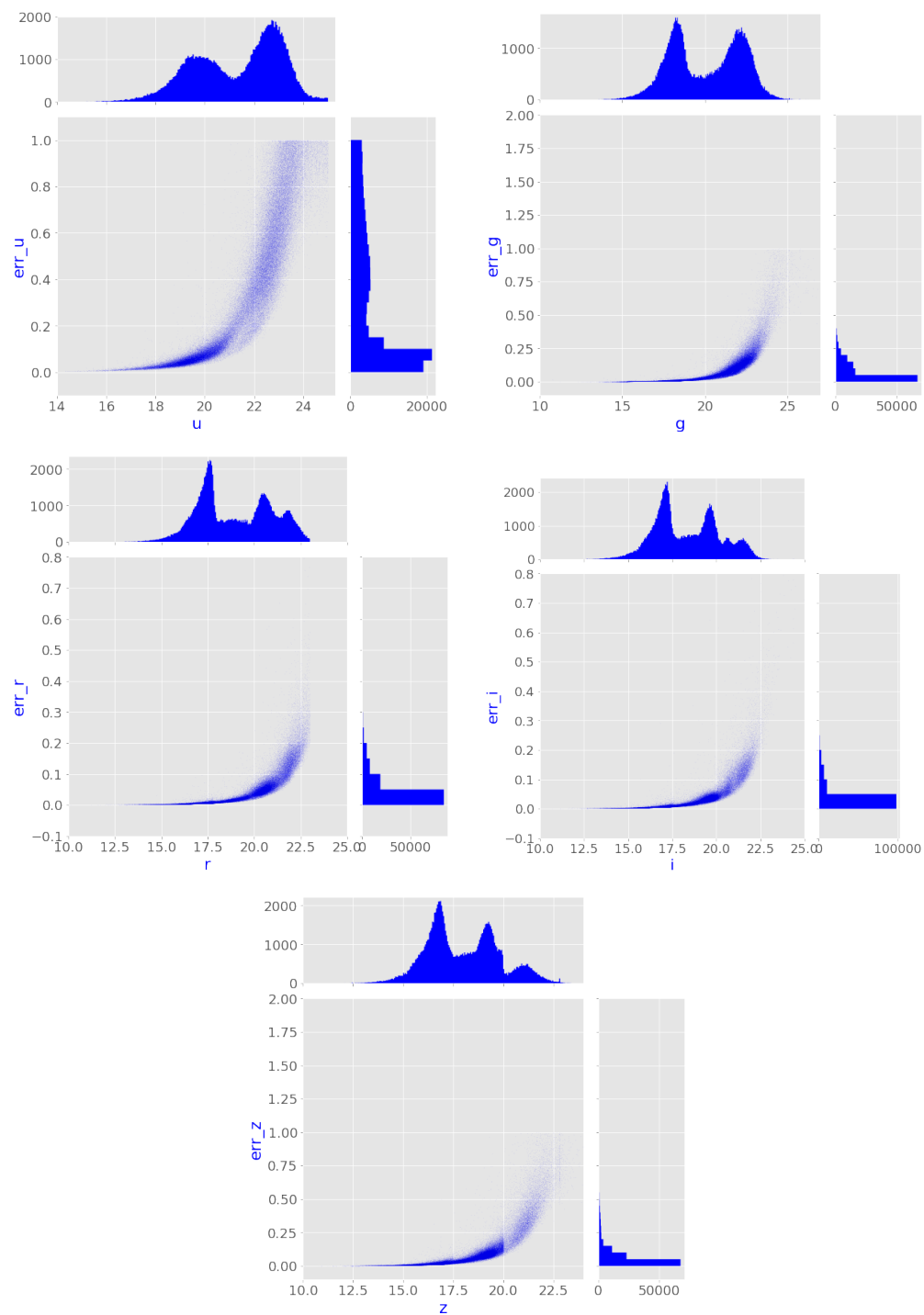


Figura 28 – Distribuição das magnitudes pelos erros em cada banda do conjunto SDSS-R16-120K.

5.3.2- COIN/Teddy

O catálogo Teddy, semelhantemente ao Happy, é disponibilizado em quatro arquivos denominados pelas letras A, B, C e D. O conjunto A corresponde ao conjunto de treinamento e validação, e os conjuntos B, C e D correspondem aos conjuntos de teste. Para facilitar a comparação de resultados em nossos experimentos, utilizamos essa mesma divisão já realizada pela equipe do projeto COIN. Tais dados, conforme descritos em [Beck et al., 2017b], se dividem em dois conjuntos maiores denominados estreito e largo, os quais foram usados para compor os subconjuntos A, B, C e D. O conjunto estreito se refere a objetos contidos em um espaço de cor 4D retangular representados por cortes e limitações nas bandas u , g , r , i e z . O conjunto largo se refere aos objetos fora do espaço pertencente ao conjunto estreito. Seguem os detalhes sobre como foram compostos cada um dos arquivos segundo [Beck et al., 2017b]:

- A: Amostra aleatoriamente capturada da amostragem espectroscópica do SDSS-DR12 que representa uma primeira parte do conjunto estreito.
- B: Amostra semelhante a de A, mas representa uma segunda parte do conjunto estreito.
- C: Amostra possui a mesma cobertura em magnitude/espaco de cor que A e B, mas selecionando os objetos do conjunto largo.
- D: Amostra possui objetos aleatórios do conjunto largo, possuindo assim uma cobertura em magnitude/espaco de cor maior do que em A e B.

Para este conjunto de dados removemos a coluna correspondente ao atributo identificador (ID) por se tratar apenas de um metadado irrelevante para os experimentos. Além disso não aplicamos nenhum tipo de tratamento, usando todos os dados contidos de forma integral. Esta amostragem possui apenas galáxias e aproximadamente 74.000 objetos de exemplos no conjunto de treinamento e validação (Conjunto A). A Tabela 9 apresenta as quantidades de registros nos conjuntos A, B, C e D desta amostragem.

A Figura 29 demonstra a matriz de correlação dos dados e notamos que para esta amostragem a matriz é mais densa comparada aos outros conjuntos (Happy Figura 33 e SDSS Figura 21). Todas as dimensões possuem considerável grau de influências tanto

Conjunto	Objetos	Finalidade
A	74.309	Treinamento e Validação
B	74.559	Teste
C	97.980	Teste
D	75.925	Teste

Tabela 9 – Características dos arquivos dos conjuntos de treinamento, validação e teste dos dados da amostragem COIN/Teddy de forma comparativa.

no valor alvo `redshift` quanto nas demais dimensões. Isto comprova a relevância de todas essas dimensões na direção de se obter um bom modelo preditivo, com a exceção da dimensão `redshiftErr` a qual excluímos dos experimentos.

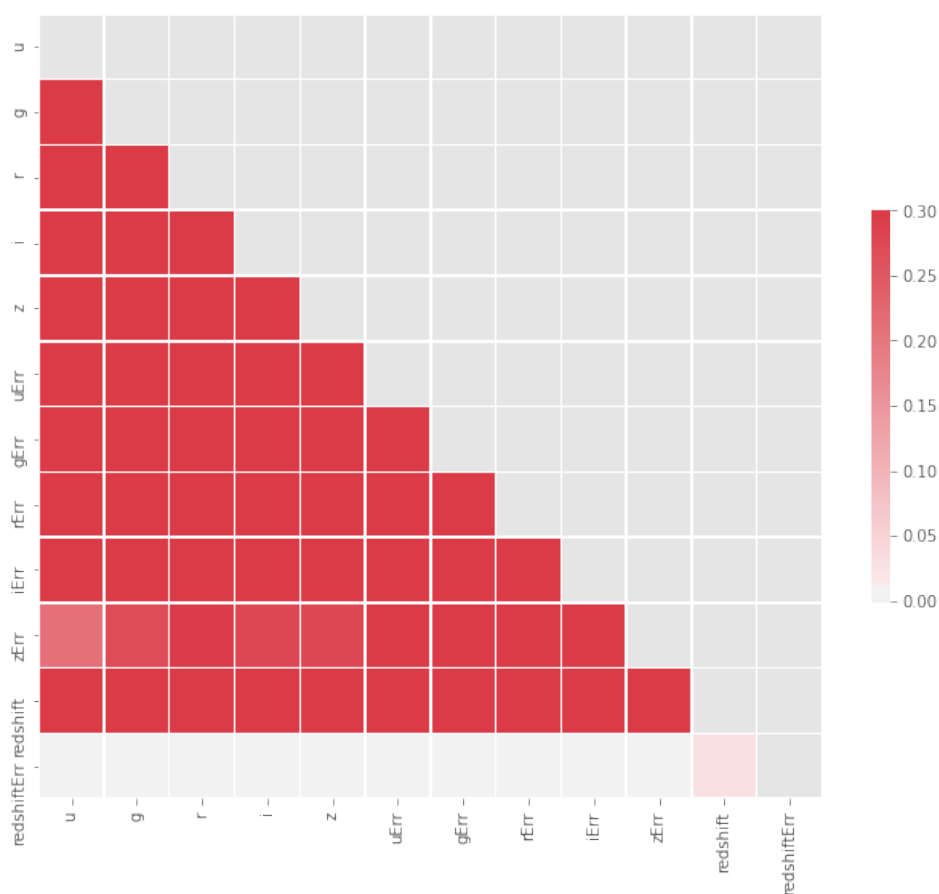


Figura 29 – Matriz de correlação do conjunto A de COIN/Teddy.

Na Figura 30 observamos a presença de valores atípicos (*outliers*) com maior concentração nas bandas `i` e `z`, o que pode ser explicado dado que as leituras destas bandas são provenientes de sinais menos brilhantes, e portanto mais erráticos.

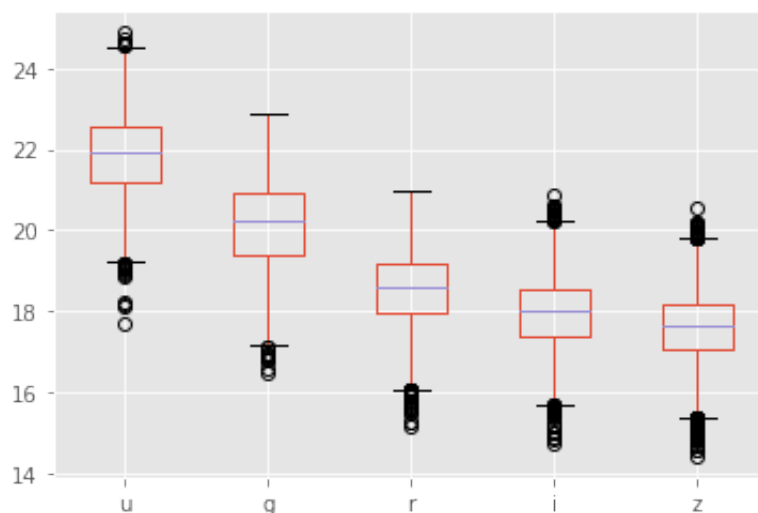


Figura 30 – Boxplot das magnitudes com os valores atípicos do conjunto A de COIN/Teddy.

Na Figura 31 observamos a presença de valores atípicos nas componentes dos erros, o qual possui maior concentração desses valores em u_{Err} . Este conjunto apresenta a menor amostragem de valores atípicos em relação a Happy (Figura 35) e SDSS-R16-120K (Figura 26).

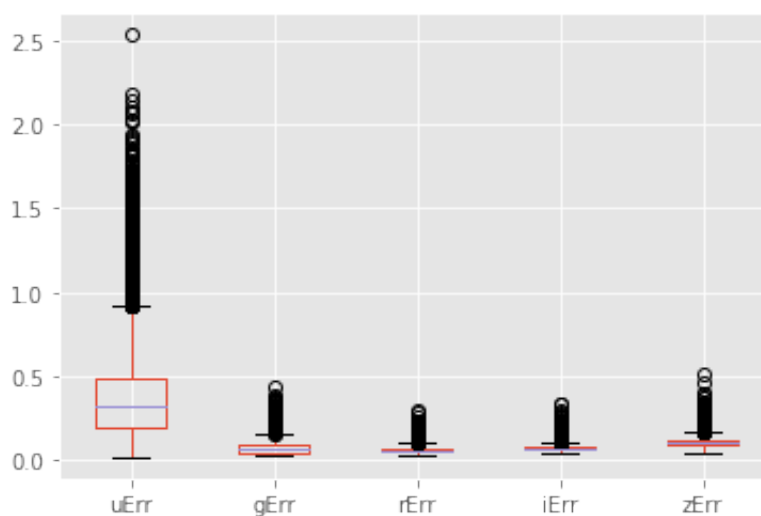


Figura 31 – Boxplot com os valores atípicos dos erros do conjunto A de COIN/Teddy.

A Figura 32 apresenta a distribuição das magnitudes pelos erros em cada uma das bandas (u, g, r, i, z). Nessa figura, podemos observar que há uma crescente dos erros em relação à magnitude: quanto menor o brilho, maior o erro em cada uma das bandas observadas. Para as bandas i e z este comportamento crescente é atenuado.

Quanto aos histogramas apresentados na Figura 32 observamos que é um

catálogo mais uniforme em relação ao de Happy (Figura 36) e SDSS-R16-120K (Figura 28) por apresentar um único pico entre todas as bandas. Observamos também que a maior concentração dos dados ocorre por volta de 0.1 dos erros e cada uma das bandas.

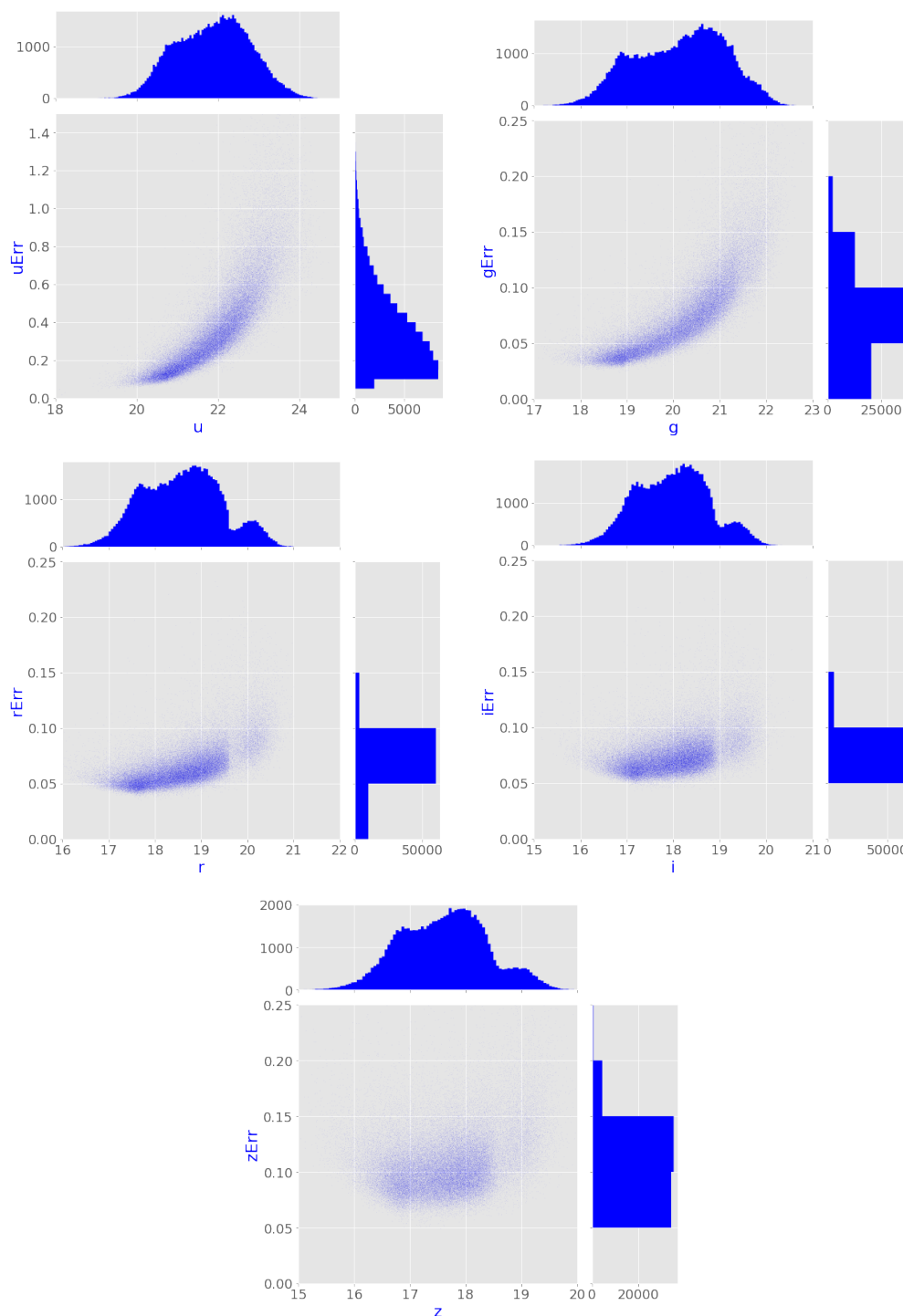


Figura 32 – Distribuições das magnitudes pelos erros em cada banda do conjunto A de COIN/Teddy.

5.3.3- COIN/Happy

O catálogo Happy, semelhantemente ao de Teddy, é disponibilizado também em quatro arquivos denominados pelas letras A, B, C e D. O conjunto A corresponde ao conjunto de treinamento e validação, e os conjuntos B, C e D correspondem aos conjuntos de teste. Para facilitar a comparação de resultados em nossos experimentos, utilizamos essa mesma divisão já realizada pela equipe do projeto COIN. Conforme descrito em Beck et al. [2017b], essas amostras mantêm as mesmas propriedades quanto aos valores de magnitude da amostra original do SDSS-DR12, porém foram misturados dados de outras origens espectroscópicas. Esta abordagem fornece uma amostra mais profunda de galáxias espectroscópicas, ao mesmo tempo em que mantém o uso da fotometria real e sistemática inerente do SDSS. Seguem os detalhes sobre cada um dos arquivos:

- A: Amostra espectroscópica semelhante ao original de SDSS-DR12 na distribuição dos objetos, selecionado aleatoriamente por meio da técnica de vizinhos próximos na banda r .
- B: Amostra semelhante a de A, ou seja, completamente representativo de A e, portanto, espelha o cenário equivalente dos exercícios tradicionais de validação de foto-z.
- C: Amostra possui a mesma gama de erros fotométricos de A e B. Possui também a mesma cobertura de cor/magnitude, mas a distribuição de erros é muito mais ponderada para erros altos. Esta ilustra uma amostra fotométrica que foi cortada para se adequar à A e B, com uma proporção maior de objetos com erros fotométricos elevados.
- D: Amostra fotométrica semelhante ao original de SDSS-DR12 na distribuição de todos os objetos, servindo como uma amostra fotométrica completa, com uma cobertura de parâmetro mais ampla e erros de medição mais elevados.

Para este conjunto de dados, também removemos a coluna correspondente ao atributo identificador (ID), e não aplicamos nenhum tipo de tratamento (i.e., usamos a totalidade dos dados obtidos). Esta amostragem possui apenas galáxias e aproximada-

mente 75.000 objetos no conjunto de treinamento e validação (Conjunto A). A Tabela 10 descreve os conjuntos A, B, C e D desta amostragem.

Conjunto	Objetos	Finalidade
A	74.950	Treinamento e Validação
B	74.900	Teste
C	60.315	Teste
D	74.642	Teste

Tabela 10 – Características dos arquivos dos conjuntos de treinamento, validação e teste dos dados da amostragem COIN/Happy de forma comparativa.

A Figura 33 apresenta a matriz de correlação dos dados. Notamos que a matriz é consideravelmente densa, resultante do fato de existir correlações positivas altas entre vários pares de magnitudes, e entre magnitudes e o desvio fotométrico (*redshift*). Apenas as dimensões *iErr* e *zErr* apresentam um menor grau de influência nesse conjunto, mas não o suficiente para desconsiderá-lo.

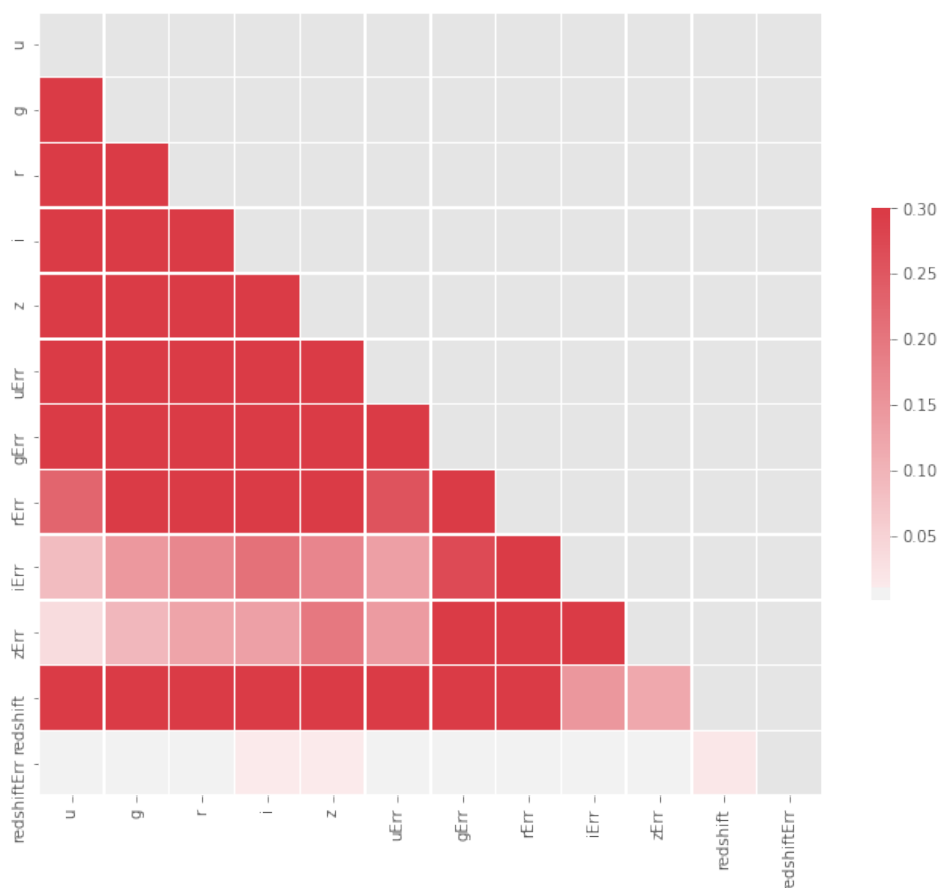


Figura 33 – Matriz de correlação do conjunto de amostras COIN/Happy.

Na Figura 34, observamos a presença de valores atípicos (*outliers*) com maior

concentração nas bandas *i* e *z*. Isso pode ser explicado considerando que as leituras destas bandas são provenientes de sinais menos brilhantes, e portanto mais erráticos. Em comparação ao conjunto COIN/Teddy na Figura 30, este é o que possui mais valores atípicos nas regiões superiores, a partir da magnitude 23 em todas as bandas, mas principalmente em *u*, *g* e *r*.

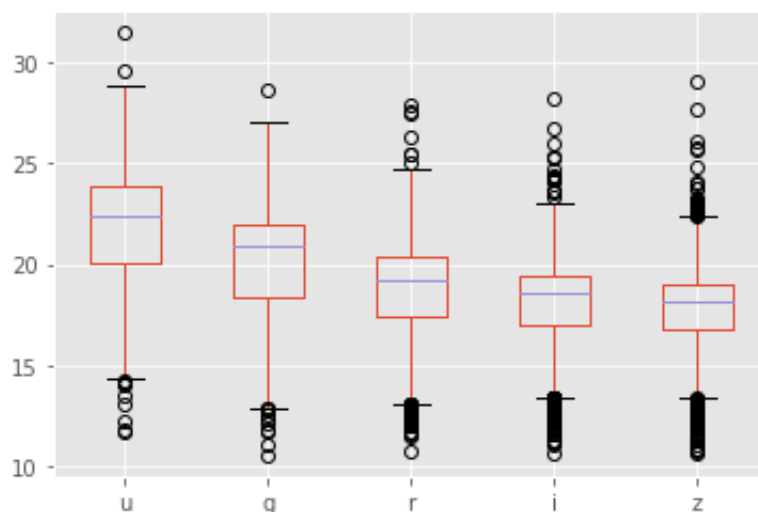


Figura 34 – Boxplot contendo os valores atípicos das magnitudes do conjunto COIN/Happy.

Na Figura 35 observamos a presença de valores atípicos nas componentes dos erros, o qual possui maior concentração desses valores em *gErr*, *iErr* e *zErr*. Estes valores podem ser explicado por conta da mistura dos dados fotométricos, os quais possuem valores mais erráticos nessas componentes tal qual visto em [Beck et al., 2017b].

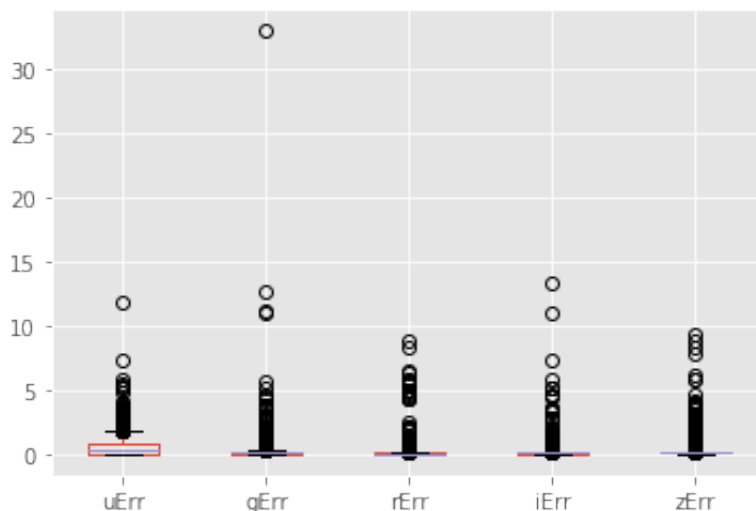


Figura 35 – Boxplot com os valores atípicos dos erros do conjunto A de COIN/Happy.

A Figura 36 apresenta a distribuição das magnitudes pelos seus respectivos erros em cada uma das bandas (u , g , r , i , z). Nessa figura, podemos observar a tendência de que quanto menor o brilho, maior o erro em cada uma das bandas observadas. Ao observar a distribuição da banda u , notamos uma característica particular: uma tendência decrescente a partir da magnitude 25, o que também é observado no conjunto SDSS-R16-120K (Figura 28). Outra particularidade observada neste conjunto são as curvas em formato côncavo nas bandas r , i e z provavelmente oriundas das leituras fotométricas, que têm um caráter mais errático, dada a natureza da captura dessas amostras, e isto estaria perturbando essas curvas neste aspecto. Algo desta natureza era de certa forma esperado por conta da proposta da amostragem Happy, que foi criada para se avaliar os impactos dos erros fotométricos na estimativa de foto- z [Beck et al., 2017b].

Quanto aos histogramas apresentados na Figura 36 observamos que há uma descontinuidade no catálogo das bandas u , g por volta da magnitude 20, e nas bandas r , i e z por volta da magnitude 17.5. Este conjunto apresenta semelhanças quanto a sua descontinuidade em relação ao de SDSS-R16-120K (Figura 28) porém um pouco mais completo. Observamos também que a maior concentração dos erros ocorrem por volta de 0.1 dos erros e cada uma das bandas.

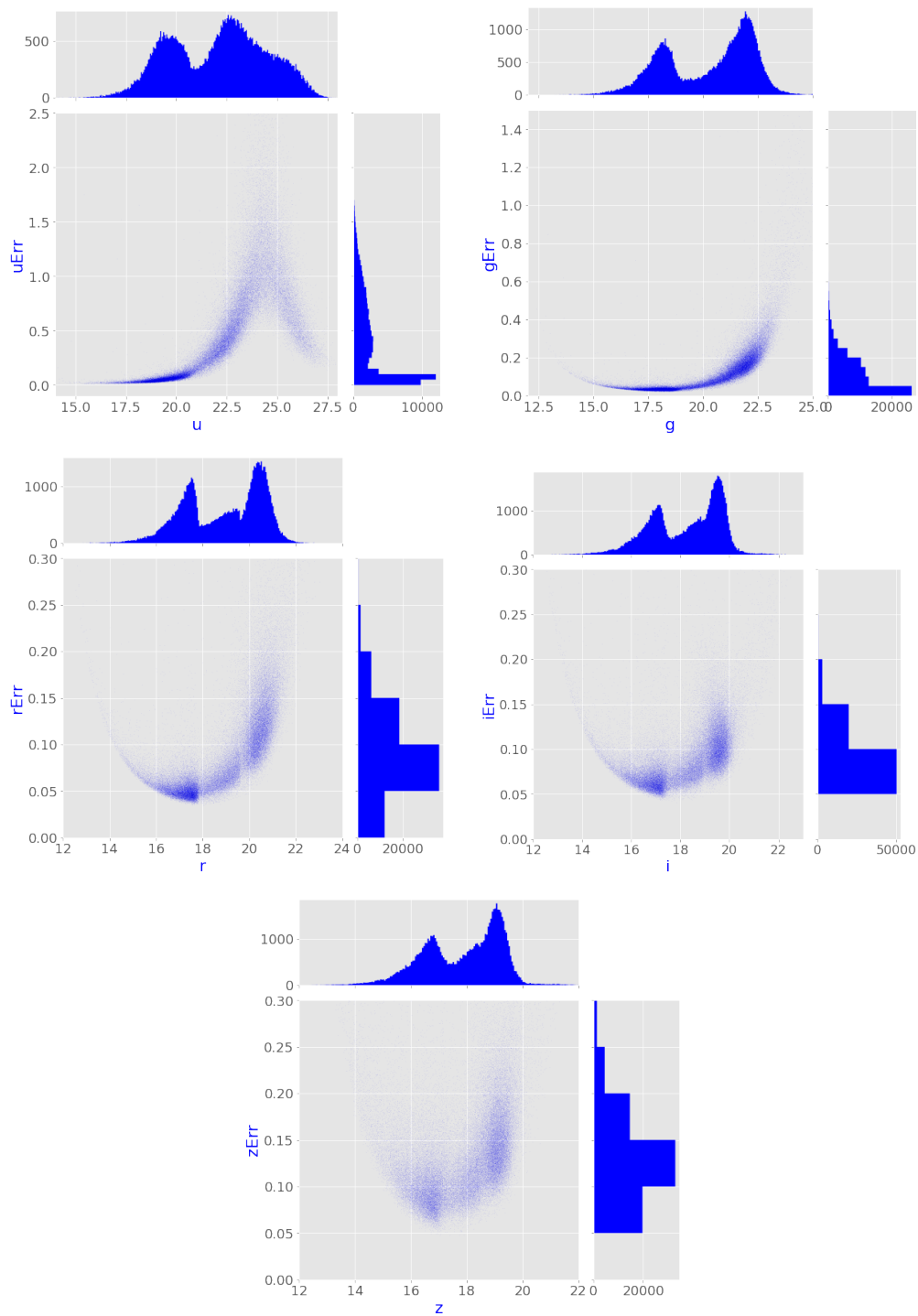


Figura 36 – Distribuição das magnitudes pelos erros em cada banda do conjunto A de COIN/Happy.

5.3.4- Divisão dos conjuntos de dados

Após a análise exploratória e a eventual remoção de valores extremos, realizamos a divisão dos conjuntos de dados disponíveis em subconjuntos para treinamento, validação e teste dos modelos, conforme descrito a seguir.

- Para o treinamento de modelos com os catálogos COIN/Teddy e COIN/Happy, realizamos a separação aleatória do conjunto A entre treinamento e validação da seguinte forma: 80% para treinamento e 20% para de validação. Os conjuntos B, C e D foram usados apenas na fase de testes dos modelos preditivos gerados.
- Para o treinamento de modelos usando o conjunto do SDSS-R16-120K, realizamos a separação aleatória entre conjunto de treinamento, validação e teste da seguinte forma: 60% para treinamento, 20% para de validação e 20% para testes.

5.4- Métricas de Avaliação

A predição de *redshift* fotométrico é um tipo particular de tarefa de regressão. Sendo assim, utilizamos a métrica do desvio médio absoluto para avaliação dos modelos construídos, mantendo-se de forma comparável aos resultados de Beck et al. [2017a]. Para análise dos resultados produzidos, também adotamos o coeficiente de determinação, e a distância (medida em erros padrão) entre o *redshift* espectroscópico (*z-spec*) e o predito (*z-photo*). Descrevemos essas medidas a seguir.

5.4.1- MAD

O desvio mediano absoluto (*median absolute deviation* - MAD) é computado conforme as Equações 16 e 17. O $z_{spec}^{(i)}$ representa os valores alvos na *i*-ésima posição do conjunto de teste, e $z_{phot}^{(i)}$ os valores preditos produzidos pelo modelo na Equação 16.

Na Equação 17, a mediana é computada por meio dos valores absolutos da coleção Δz_{norm} produzida na Equação 16.

$$\Delta z_{norm}^{(i)} = \frac{(z_{spec}^{(i)} - z_{phot}^{(i)})}{(1 + z_{spec}^{(i)})} \quad (16)$$

$$MAD(\Delta z_{norm}) = \tilde{\delta}(|\Delta z_{norm}|) \quad (17)$$

Esta estatística dá uma intuição sobre os resíduos entre os valores alvo de *redshift* espectroscópico e os produzidos de *redshift* fotométrico. Quanto mais próximos de zero forem os resíduos, melhores serão os resultados produzidos pelo modelo.

5.4.2- Taxa de valores extremos (*outliers*)

A medida RMSE (Seção A.1) pode ser interpretada como o desvio padrão dos erros gerados pelo modelo de regressão, conforme a Equação 18. Nesse caso, esse valor é denominado erro padrão da estimativa (*standard error of the estimate*).

$$\sigma_{est} = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2} \quad (18)$$

5.5- Regressores usados nos experimentos

Para fins de comparação com o método proposto nesta dissertação, usamos os seguintes regressores de *redshift* nos experimentos: XGBoost (Seção 2.2.3), redes neurais artificiais MLP (Seção 2.2.1).

Para cada método usado para comparação, realizamos o treinamento em dois cenários distintos. No primeiro, usamos como dimensões ou características do treinamento as componentes de magnitudes u, g, r, i, z. No segundo cenário, usamos os valores de magnitude e os valores dos erros respectivos. O uso desses dois cenários (ou

seja, experimentos com ugriz e com ugriz+erros) tem o propósito de permitir verificarmos o impacto dos erros na predição. Considerando a pergunta de pesquisa apresentada na Seção 1.3, queremos verificar se ignorar essas componentes de erros, ou considerá-las como meras características, seria o melhor caminho para construir um modelo preditivo mais eficiente.

Desenvolvemos a lista de abreviações a seguir para facilitar as referências às diferentes variações de regressores usados nos experimentos.

- RNA05: Rede neural artificial MLP treinada apenas com as componentes de magnitudes.
- RNA10: Rede neural artificial MLP treinada com as componentes de magnitudes e erros.
- XGB05: XGBoost treinado apenas com as componentes de magnitudes.
- XGB10: XGBoost treinado com as componentes de magnitudes e erros.

Para o método proposto nesta dissertação (Capítulo 4), também investigamos múltiplas variações de regressores, conforme lista a seguir.

- RNA-RAD-Inv: Rede neural artificial treinada com camada inspirada em Dropout por meio de árvores de decisão para estimativa dos erros.
- RNA-RRI-Inv: Rede neural artificial treinada com camada inspirada em Dropout por meio de regressão isotônica para estimativa dos erros.

Também investigamos a aplicação de normalização (*scaling*) sobre os valores de magnitudes para alguns experimentos. Nenhuma normalização foi aplicada sobre os valores alvos ou sobre os valores de erros referentes às magnitudes.

Cada variação foi treinada por 10 vezes para assegurar baixa variância nas estimativas de predição computadas e apresentadas na Seção 5.6. Em seguida, computamos a média e o desvio-padrão considerando as dez rodadas.

Para todos os modelos preditivos treinados, baseados em redes neurais ou não, computamos as métricas descritas na Seção 5.4. Para os modelos treinados com o algoritmo XGBoost (XGB), usamos a implementação `GradientBoostingRegressor` do Scikit-Learn⁵, e configuramos os seus hiperparâmetros como se segue na Tabela 11.

⁵<https://scikit-learn.org/stable/>

O valor $\$epochs$ conforme as épocas usadas nos experimentos foi de 3.000, e o valor representado por $\$f$ representa 20% do valor de $\$epochs$. O atributo ls do hiperparâmetro $loss$ se refere à regressão de mínimos quadrados (*least squares*), e é o valor padrão.

Hiperparâmetro	Valor
<code>n_estimators</code>	$\$epochs$
<code>max_depth</code>	8
<code>min_samples_split</code>	5
<code>validation_fraction</code>	0.2
<code>n_iter_no_change</code>	$\$f$
<code>learning_rate</code>	0.001
<code>loss</code>	ls

Tabela 11 – Hiperparâmetros usados no Regressor XGBoost.

5.5.1- Configurações relacionadas às redes neurais artificiais

Para todas as arquiteturas de Redes Neurais Artificiais (RNA), usamos duas camadas ocultas. Para definir a quantidade de unidades em cada camada oculta, usamos a regra de Huang [2003] que visa selecionar as quantidades de neurônios em cada camada oculta, especialmente para redes compostas por duas camadas ocultas, tais como as usadas nos experimentos executados neste trabalho. Huang [2003] propõe usar a Equação 19 e a Equação 20 para determinar as quantidades de neurônios na primeira e segunda camadas ocultas, respectivamente. Nessas equações, m denota a quantidade de neurônios da camada de saída, e N representa a quantidade de exemplos contida no conjunto de treinamento. A única adaptação que fizemos nas equações foi a de arredondar o valor com a finalidade de obtermos o número inteiro mais próximo.

$$h_1 = \left\lceil \sqrt{(m+2)N} + 2\sqrt{\frac{N}{(m+2)}} \right\rceil \quad (19)$$

$$h_2 = \left\lceil m\sqrt{\frac{N}{(m+2)}} \right\rceil \quad (20)$$

Posto que todos os conjuntos de treinamentos usados possuem aproximadamente 75000 amostras, e a quantidade de neurônios na camada de saída é 1, aplicamos as

equações acima com os valores $m = 1$ e $N = 75000$. Como resultado, obtivemos 791 neurônios para a primeira camada oculta, e 158 neurônios para a segunda.

Em todos os modelos de redes MLP treinados, utilizamos a função de ativação ReLU (*Rectified Linear Unit*), apresentada na Equação 21, em que w representa o vetor de pesos da unidade, e x é o vetor que representa o padrão de entrada.

$$\text{ReLU}(x) = \max(0, w^T \cdot x) \quad (21)$$

Para todos os modelos de MLP treinados, usamos o otimizador Adam, que é uma variante do método gradiente descendente estocástico. Esse otimizador é computacionalmente eficiente com pouco consumo de memória, e é adequado para problemas que são grandes em termos de dados ou parâmetros [Kingma and Ba, 2014]. Para a função de custo selecionamos a função de erro médio quadrático (MSE). Configuramos a taxa de aprendizado (*learning rate*) inicial em 0.001 (esse é o valor padrão do otimizador Adam). Para a inicialização dos pesos em cada camada, usamos a implementação padrão que a biblioteca *Keras*⁶ fornece por meio da classe *Dense*⁷. Esta inicia seus pesos com o vetor de vieses zerados, por meio do parâmetro `bias_initializer='zeros'`, e usa para o kernel da matriz de pesos o método inicializador uniforme Glorot, também chamado de inicializador uniforme Xavier [Glorot and Bengio, 2010] por meio do parâmetro `kernel_initializer='glorot_uniform'`. Configuramos o hiperparâmetro de tamanho do lote `batch_size` em 32, padrão da classe *Model*⁸ da biblioteca *Keras*.

5.6- Resultados

Organizamos os resultados dos experimentos referentes ao conjunto Teddy nas Tabelas 16 e 13, e nas Figuras 38 e 41. Happy nas Tabelas 17 e 14, e nas Figuras 39 e 42. Amostragem do SDSS-R16-120K nas Tabelas 15 e 12, e nas Figuras 37 e 40. Podemos observar o desempenho superior da *baseline* RNA10 sobre todos os conjuntos (COIN/Teddy, COIN/Happy e SDSS-R16-120K). O modelo apresentou o menor valor de

⁶<https://keras.io/>

⁷https://www.tensorflow.org/api_docs/python/tf/keras/layers/Dense

⁸https://www.tensorflow.org/api_docs/python/tf/keras/Model

MAD, vide Tabelas 12, 13 e 14. Nossos modelos propostos representados pela RNA-RAD-Inv e RNA-RRI-Inv obtiveram resultados comparáveis aos *baselines* RNA05 e RNA10, ficando por uma diferença pouco expressiva. Incluir os erros como características no treinamento colaborou para um melhor desempenho do estimador, ao se tratar de RNA10. Isto é apoiado pelas observações nas matrizes de correlação (Figuras 21, 29 e 33), uma vez em que ao analisar os conjuntos de dados, todos mostraram relação entre o *redshift* e os erros. Um resultado similar se obteve no trabalho de D’Isanto and Polsterer [2018], onde também os modelos em que tiveram os erros incluídos no treinamento obtiveram resultado surpreendente melhor. Não há relação física aparente entre o *redshift* dos objetos e os erros relativos dos sinais de magnitude medidos pelo sistema fotométrico do SDSS conforme D’Isanto and Polsterer [2018] também discutem. Porém, os erros são normalmente maiores para objetos de maior magnitude, significando que esses são objetos fisicamente mais distantes.

Método	MAD ($\times 10^{-2}$)	Tempo (min)
RNA-RAD-Inv	1.87 ± 0.16	67
RNA-RRI-Inv	1.88 ± 0.17	65
RNA05	1.73 ± 0.00	62
RNA10	1.70 ± 0.05	64
XGB05	2.24 ± 0.00	16
XGB10	2.22 ± 0.00	31

Tabela 12 – Resultados após 3.000 épocas, com 791 unidades na primeira camada oculta e 158 unidades na segunda camada oculta, usando o conjunto SDSS-R16-120k.

Método	MAD ($\times 10^{-2}$)	Tempo (min)
B		
RNA-RAD-Inv	1.20 \pm 0.04	64
RNA-RRI-Inv	1.18 \pm 0.04	63
RNA05	1.26 \pm 0.09	62
RNA10	1.13\pm0.01	60
XGB05	1.31 \pm 0.00	12
XGB10	1.31 \pm 0.00	24
C		
RNA-RAD-Inv	1.20 \pm 0.05	64
RNA-RRI-Inv	1.18 \pm 0.06	63
RNA05	1.28 \pm 0.12	62
RNA10	1.13\pm0.01	60
XGB05	1.36 \pm 0.00	12
XGB10	1.37 \pm 0.00	24
D		
RNA-RAD-Inv	2.76 \pm 1.06	64
RNA-RRI-Inv	2.81 \pm 1.02	63
RNA05	2.88 \pm 0.51	62
RNA10	2.49\pm0.26	60
XGB05	5.02 \pm 0.00	12
XGB10	4.91 \pm 0.00	24

Tabela 13 – Resultados após 3.000 épocas, com 791 unidades na primeira camada oculta e 158 unidades na segunda camada oculta, usando o conjunto COIN/Teddy A para treinamento/validação e B, C e D para teste.

Método	MAD ($\times 10^{-2}$)	Tempo (min)
B		
RNA-RAD-Inv	1.57 \pm 0.11	66
RNA-RRI-Inv	1.52 \pm 0.10	65
RNA05	1.46 \pm 0.02	62
RNA10	1.42\pm0.02	62
XGB05	1.77 \pm 0.00	12
XGB10	1.81 \pm 0.00	24
C		
RNA-RAD-Inv	3.71 \pm 0.15	66
RNA-RRI-Inv	3.63 \pm 0.13	65
RNA05	3.53 \pm 0.00	62
RNA10	3.37\pm0.05	62
XGB05	4.19 \pm 0.00	12
XGB10	4.05 \pm 0.00	24
D		
RNA-RAD-Inv	5.64 \pm 0.12	66
RNA-RRI-Inv	5.59 \pm 0.09	65
RNA05	5.54 \pm 0.07	62
RNA10	5.21\pm0.11	62
XGB05	5.94 \pm 0.00	12
XGB10	5.82 \pm 0.00	24

Tabela 14 – Resultados após 3.000 épocas, sem uso de *early stopping*, com 791 unidades na primeira camada oculta e 158 unidades na segunda camada oculta, usando o conjunto COIN/Happy A para treinamento/validação e B, C e D para teste.

Ainda em relação aos resultados das Tabelas 12, 13 e 14 notamos que o tempo de processamento dos treinamentos dos modelos de regressão XGBoost se mostraram mais rápidos que os demais, em especial o do XGB05. Porém, em comparação ao resultado observado no MAD, acreditamos que não compense em usar esse modelo apenas por ser mais rápido em detrimento de seu desempenho preditivo inferior. Não incluímos o

desvio padrão nas leituras de tempo por estas se mostrarem pouco relevante. Em todas as execuções, os tempos de treinamento se mostraram muito constantes dentro de um mesmo método. Execuções com os tempos praticamente cravados.

As demais *baselines* (XGB05 e XGB10) foram inferiores aos nossos modelos propostos, as quais se mostraram com inferior poder preditivo. Acreditamos que a regularização aplicada trouxe melhorias pelo fato de criarmos mais exemplos para a rede, e de suprimirmos alguns exemplos menos confiáveis durante a fase de treinamento. A regularização que aplicamos em nossa proposta tem como objetivo manter os exemplos que tenham os menores erros esperados. Quanto maior for o erro esperado em uma determinada banda, maior é a probabilidade de ele ser desconsiderado. A princípio a nossa intuição era de que a abordagem proposta venceria em algum dos cenários, dadas as crescentes das curvas observadas nos sinais de magnitudes pelos erros, como demonstrada nas Figuras 28, 32 e 36. A justificativa provável para a derrota da abordagem proposta estaria relacionada a não inclusão dos erros para o treinamento dos modelos. Ao comparar com o desempenho das *baselines* RNA05 e XGB05 (Tabelas 13 e 14), onde apenas considerou as magnitudes como características no treinamento, também não obtiveram um resultado superior. Porém, considerando o resultado conforme a Tabela 12, há uma exceção para RNA05 que demonstrou surpreendentemente um desempenho melhor, ficando abaixo apenas da RNA10. Acreditamos que a aleatoriedade dessa amostragem possa ter desfavorecido no desempenho do método proposto de alguma forma. Uma outra justificativa para a derrota do modelo proposto seria devido ao fato de desligarmos os sinais de magnitudes, dada a aplicação do *dropout*. Se no lugar de desligar simplesmente os sinais de magnitudes em algumas das bandas os substituísse pelo erro esperado, provavelmente poderíamos obter um melhor poder preditivo dos modelos.

Uma outra consideração sobre os regressores XGBoost, é que foram os que apresentaram menor desvio padrão em seus resultados para todas as métricas obtidas neste trabalho. Apesar de seu início com valores aleatórios ao longos das 10 rodadas submetidas, todos os modelos convergiram para um mesmo mínimo local. Outro destaque para eles podemos observar nas Tabelas 19 e 20, onde obtiveram melhor desempenho na amostra D, tanto para COIN/Teddy quanto para COIN/Happy, considerando as métricas de MSE, RMSE e R2 (Seção A.1).

As Tabelas 15, 16 e 17 contém as métricas relativas às distâncias em erros padrão

(Seção 5.4) de cada variante de modelo explorada durante os experimentos. Podemos comprovar que a grande maioria dos modelos obteve uma precisão em até 1 desvio padrão em mais de 90% das saídas produzidas. Os modelos da Regressão XGBoost obtiveram resultados inferiores para todos os conjuntos. Alcançamos precisões maiores para os conjuntos do Happy com mais de 95% dos resultados em até 1 desvio padrão, incluindo todos os modelos. Há um destaque para melhor precisão da *baseline* RNA10 no conjunto COIN/Teddy D, o qual obteve 99.26% dos resultados em até 1 desvio padrão. Esperávamos que os resultados do conjunto Teddy se saíssem melhor do que os do Happy por conter as leituras dos erros de forma mais confiável, porém os dados do Happy possuem leituras espectroscópicas de melhor qualidade que acabaram influenciando nesse resultado, conforme também observado em Beck et al. [2017b].

Método	0 - 1	1 - 2	2 - 3	3 - 4	4 +
RNA-RAD-Inv	96.55	1.38	0.69	0.33	1.04
RNA-RRI-Inv	96.55	1.38	0.69	0.33	1.04
RNA05	92.21	3.87	1.7	0.78	1.43
RNA10	96.8	1.23	0.61	0.27	1.08
XGB05	96.29	1.53	0.71	0.37	1.11
XGB10	96.47	1.41	0.71	0.38	1.03

Tabela 15 – Taxas de outliers, medidos em quantidades de erros padrão dos resultados das predições realizadas pelos modelos treinados com 3.000 épocas para o conjunto SDSS-R16-120K. A coluna de método se refere ao modelo produzido, e as colunas de intervalos se referem às faixas de desvio padrão onde os resultados produzidos se concentraram.

	Método	0 - 1	1 - 2	2 - 3	3 - 4	4 +
B	RNA-RAD-Inv	94.15	3.31	1.01	0.53	1.0
	RNA-RRI-Inv	94.15	3.31	1.01	0.53	1.0
	RNA05	93.89	3.47	1.1	0.51	1.03
	RNA10	94.38	3.2	1.0	0.48	0.94
	XGB05	93.35	3.85	1.15	0.62	1.04
	XGB10	93.28	3.84	1.2	0.63	1.05
C	Método	0 - 1	1 - 2	2 - 3	3 - 4	4 +
	RNA-RAD-Inv	94.79	2.81	0.96	0.51	0.93
	RNA-RRI-Inv	94.77	2.84	0.94	0.5	0.94
	RNA05	94.5	2.97	1.06	0.54	0.94
	RNA10	94.96	2.75	0.96	0.45	0.89
	XGB05	94.31	3.11	1.13	0.58	0.87
XGB10	93.96	3.35	1.17	0.57	0.95	
D	Método	0 - 1	1 - 2	2 - 3	3 - 4	4 +
	RNA-RAD-Inv	97.08	1.73	0.48	0.21	0.5
	RNA-RRI-Inv	96.66	2.0	0.56	0.23	0.55
	RNA05	98.9	0.57	0.18	0.13	0.23
	RNA10	99.26	0.36	0.12	0.07	0.19
	XGB05	90.88	6.0	1.53	0.64	0.94
XGB10	88.99	7.1	2.01	0.82	1.08	

Tabela 16 – Taxas de *outliers* medidos em quantidades de erros padrão dos resultados. Modelos treinados com 3.000 épocas na amostragem A, e testados com as amostragens B, C e D do conjunto COIN/Teddy. A coluna de método se refere ao modelo produzido, e as colunas de intervalos se referem às faixas de desvio padrão onde os resultados produzidos se concentraram.

	Método	0 - 1	1 - 2	2 - 3	3 - 4	4 +
B	RNA-RAD-Inv	97.79	1.08	0.36	0.21	0.55
	RNA-RRI-Inv	97.86	1.03	0.39	0.2	0.52
	RNA05	97.86	1.04	0.38	0.18	0.54
	RNA10	98.11	0.91	0.35	0.18	0.45
	XGB05	97.21	1.44	0.5	0.27	0.58
	XGB10	97.14	1.56	0.47	0.26	0.56
	Método	0 - 1	1 - 2	2 - 3	3 - 4	4 +
C	RNA-RAD-Inv	96.96	1.86	0.59	0.28	0.32
	RNA-RRI-Inv	97.08	1.75	0.51	0.3	0.37
	RNA05	97.05	1.79	0.55	0.26	0.35
	RNA10	97.28	1.63	0.5	0.26	0.33
	XGB05	96.45	2.28	0.61	0.31	0.34
	XGB10	96.41	2.37	0.61	0.29	0.33
	Método	0 - 1	1 - 2	2 - 3	3 - 4	4 +
D	RNA-RAD-Inv	95.61	2.93	0.79	0.3	0.37
	RNA-RRI-Inv	95.56	2.92	0.81	0.33	0.37
	RNA05	95.49	2.89	0.88	0.36	0.38
	RNA10	96.82	2.17	0.55	0.22	0.24
	XGB05	95.15	3.33	0.92	0.28	0.31
	XGB10	95.1	3.4	0.91	0.31	0.28

Tabela 17 – Taxas de outliers medidos em quantidades de erros padrão dos resultados. Modelos treinados com 3.000 épocas na amostragem A, e testados com as amostragens B, C e D do conjunto COIN/Happy. A coluna de método se refere ao modelo produzido, e as colunas de intervalos se referem às faixas de desvio padrão onde os resultados produzidos se concentraram.

As Figuras 37, 38 e 39 são referentes aos gráficos de z -spec (alvo) por z -phot (predito) produzido de cada modelo. Obtivemos resultados comparáveis ao trabalho de Beck et al. [2017b], que demonstra um resultado mais errático para as amostras D, tanto em COIN/Teddy como em COIN/Happy em todos os modelos analisados. Para B e C dos dois conjuntos COIN, os resultados são bem próximos, com concentração de *redshift* entre $0.4 < z < 0.6$. Já em COIN/Happy essa concentração aumenta para *redshift* entre $0 < z < 0.8$. Ambos os resultados de COIN seguem crescente angular por volta dos 45 graus, como constatado também em Beck et al. [2017b]. Nos resultados do SDSS-

R16-120K temos resultados também consistentes, porém com uma concentração ainda maior, com *redshift* entre $0 < z < 1$. Porém a densidade maior ocorre entre $0 < z < 0.2$ analisando o resultado de todos os modelos explorados neste conjunto. Acreditamos que esse resultado menos preciso seja por conta de não possuir vieses de cores na amostra, dado o fator aleatório durante a amostragem, conforme já demonstrado na Seção 5.3.1.

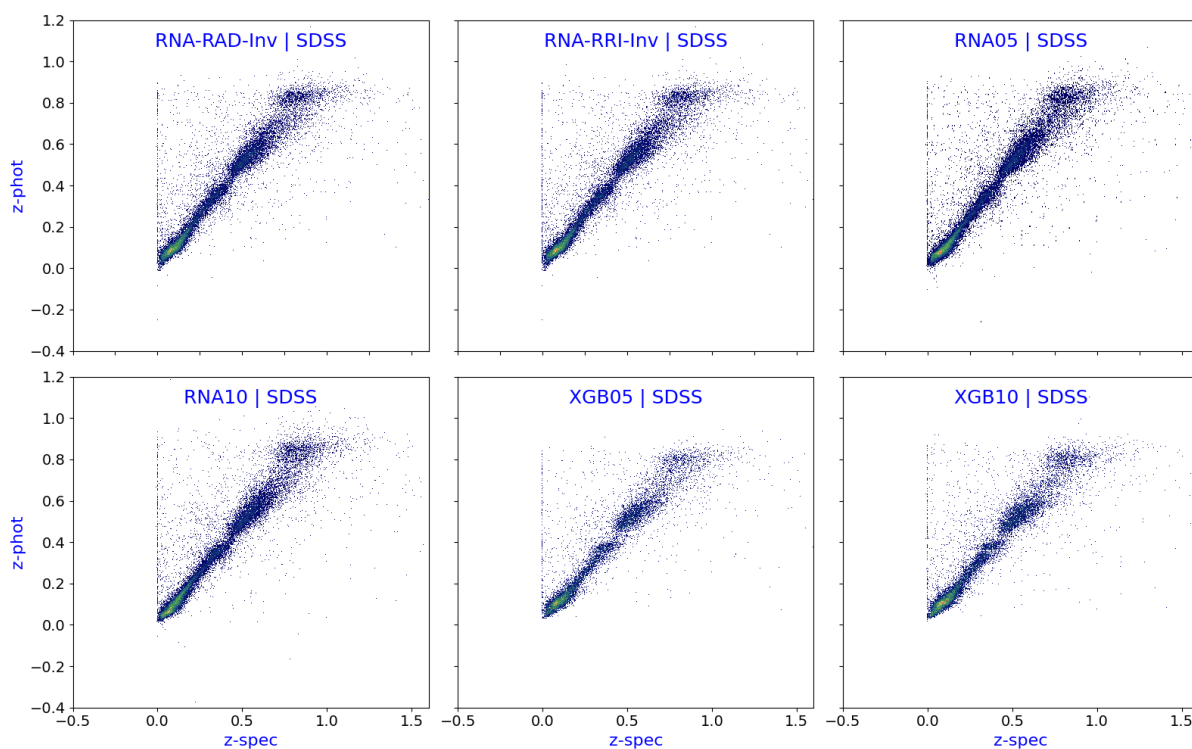


Figura 37 – Z-Spec e Z-Phot por modelo. Todos treinados por meio do conjunto SDSS-R16-120k com 3.000 épocas.

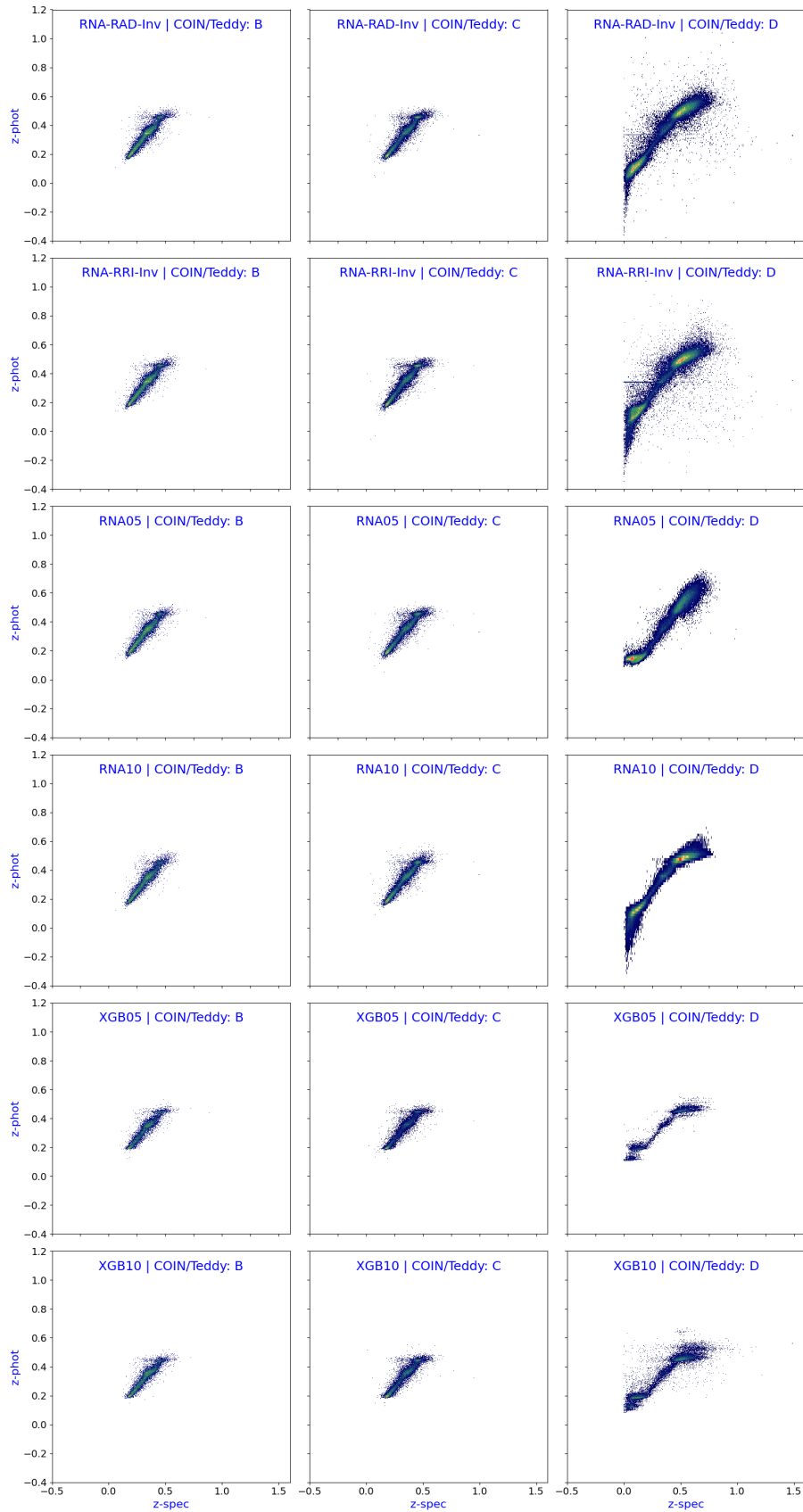


Figura 38 – Z-Spec e Z-Phot por modelo. Todos treinados por meio do conjunto COIN/-Teddy com 3.000 épocas.

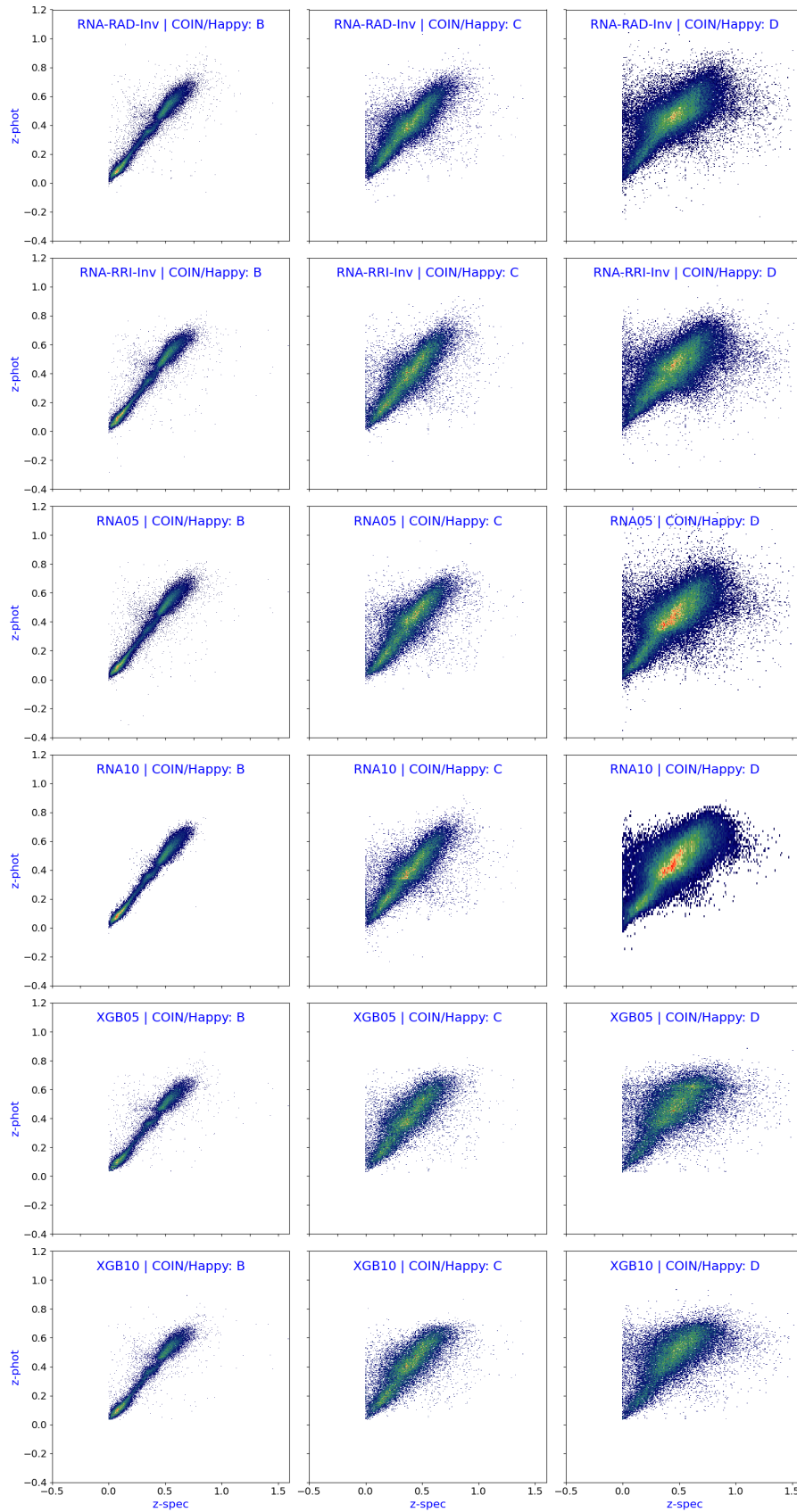


Figura 39 – Z-Spec e Z-Phot por modelo. Todos treinados por meio do conjunto COIN/Happy com 3.000 épocas.

As Figuras 40, 41 e 42 são referentes aos gráficos de z -spec (alvo) pelos resíduos (MAD) produzido de cada modelo. As faixas dos z -spec se mantiveram consistentes aos resultados analisados de z -spec por z -phot. Neste, os resultados dos modelos que mantiveram os resíduos (MAD) por volta de zero foram os que possuíram melhor desempenho preditivo. Comparando os modelos propostos com a *baseline* RNA10, notamos pouca diferença. Em RNA10 notamos uma sutil diferença nos resultados, sendo eles mais precisos dada a região em verde mais iluminada, e na região em azul menos carregada.

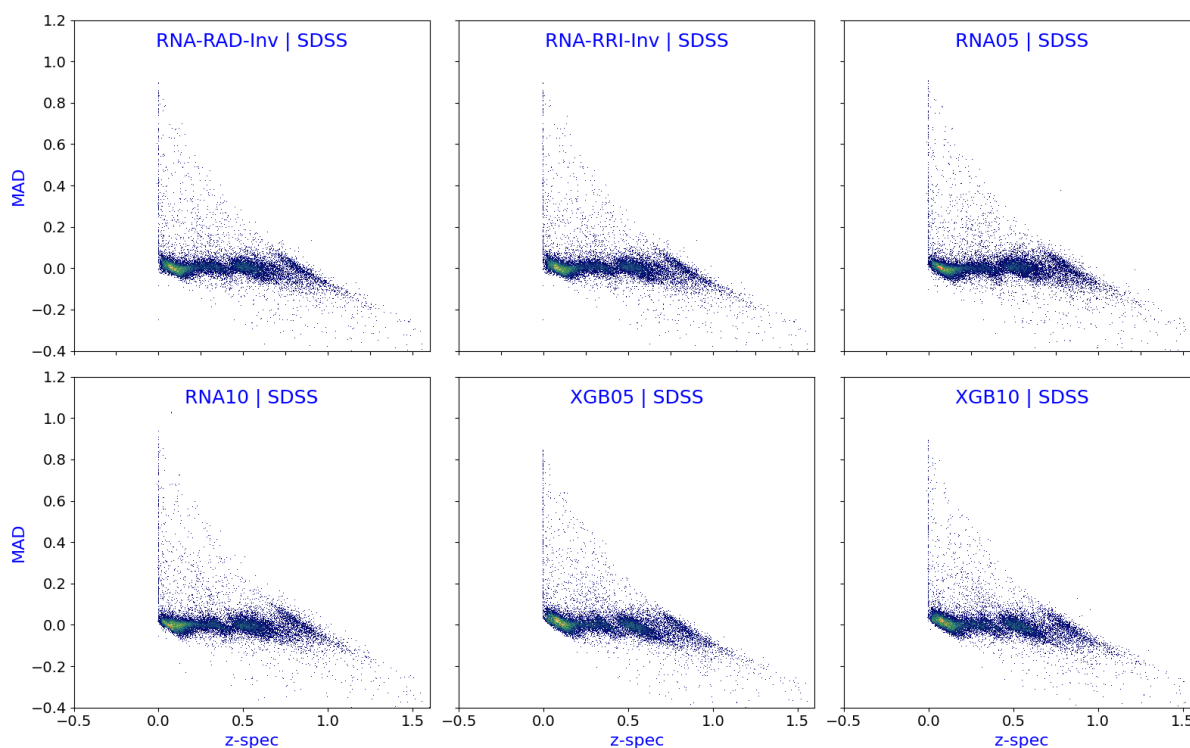


Figura 40 – Resíduos por modelo. Todos treinados por meio do conjunto SDSS-R16-120k com 3.000 épocas.

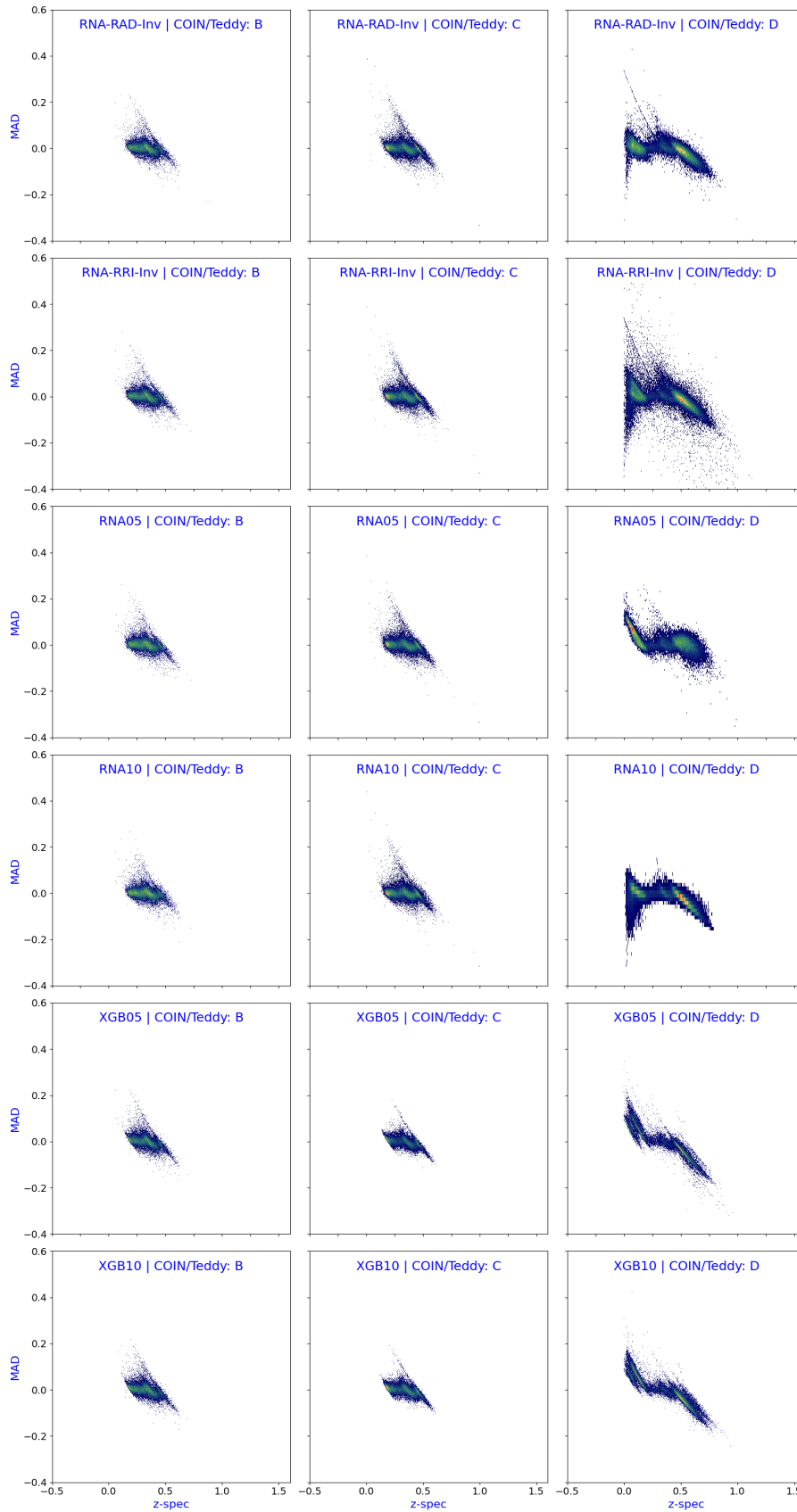


Figura 41 – Resíduos por modelo. Todos treinados por meio do conjunto COIN/Teddy com 3.000 épocas.

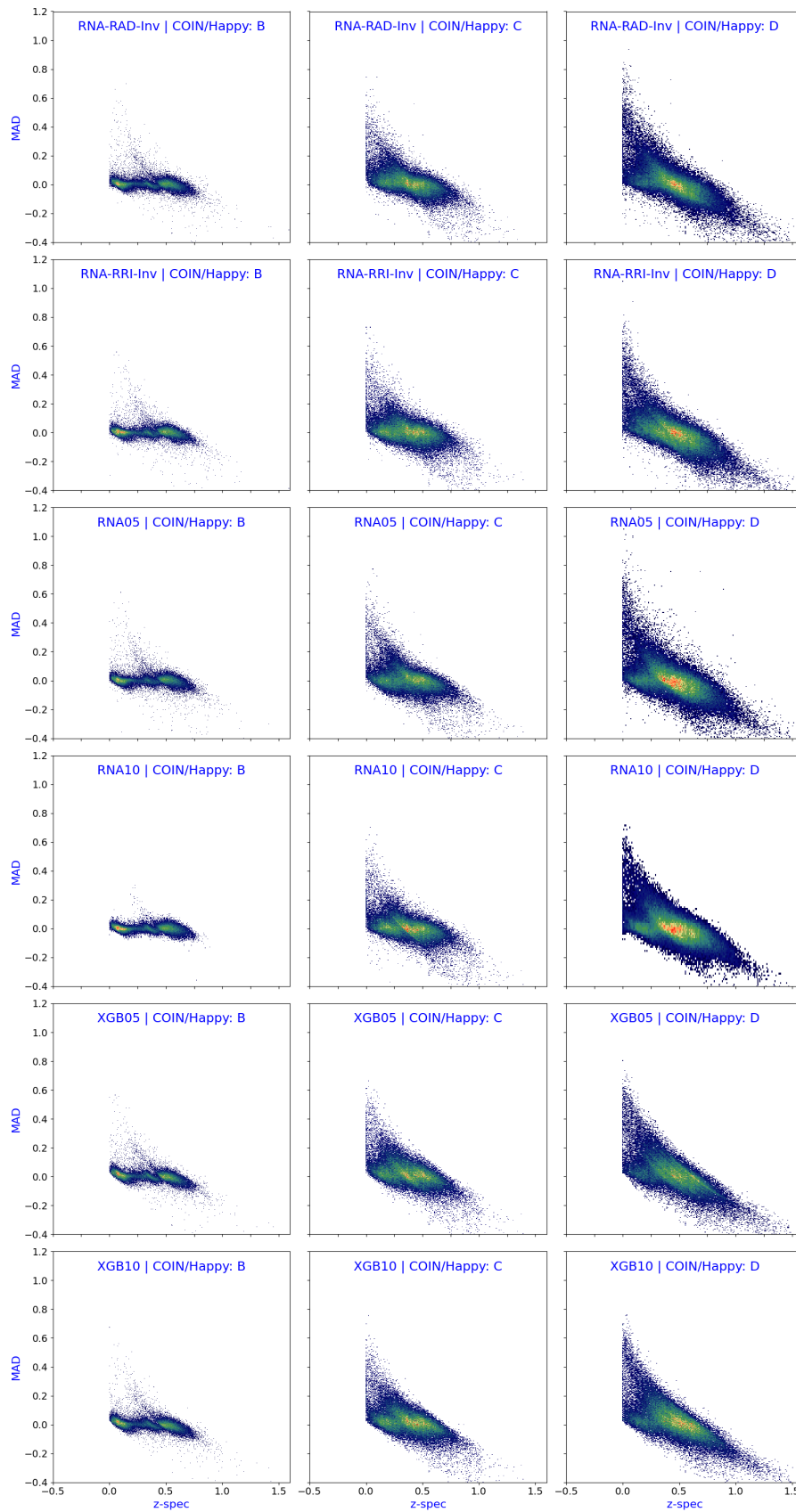


Figura 42 – Resíduos por modelo. Todos treinados por meio do conjunto COIN/Happy com 3.000 épocas.

6- Conclusões

Este capítulo apresenta as conclusões desta dissertação. Na Seção 6.1, apresentamos uma análise retrospectiva do trabalho realizado. Na Seção 6.2, apresentamos possíveis temas para trabalhos futuros.

6.1- Análise Retrospectiva

Nesta dissertação, investigamos o uso de métodos de aprendizado de máquina para gerar modelos de predição de *redshifts* fotométricos. A ideia original consistiu em utilizar as medidas de erro relacionadas às magnitudes para, durante o treinamento, calibrar a confiança do modelo nas observações de magnitude correspondentes. Intuitivamente, se uma medida de magnitude está associada a uma estimativa de erro grande, o modelo preditivo deveria dar menos importância a esse valor de magnitude durante a computação da estimativa. Porém esta abordagem não se mostrou tão eficiente quanto utilizar as medidas dos erros como características juntamente com as magnitudes. Apesar de nossa abordagem conseguir resultados comparáveis, nossa hipótese originalmente levantada não foi confirmada durante os experimentos. Em teoria, não há justificativa do ponto de vista físico, para incluir os erros durante a construção de um modelo preditivo, posto que não há nenhuma relação física aparente entre o *redshift* de uma galáxia e os erros relativos das magnitudes medidas para ela. Entretanto, objetos mais fracos (de menor brilho), que são fisicamente mais distantes, possuem erros de magnitude maior, e de certa forma essas informações colaboram na eficiência do modelo preditivo, como também observado em D'Isanto and Polsterer [2018].

Uma possível explicação para o insucesso da técnica proposta é a taxa de exemplos do conjunto de treinamento que são alterados. Fizemos alguns experimentos computacionais e constatamos que a taxa de alteração de exemplos gira em torno de 40%. Entretanto, o próprio artigo original da técnica de *Dropout* menciona que, embora essa técnica possa ser aplicada aos dados de entrada (assim como fazemos nesta dissertação),

a taxa recomendada deve ser de aproximadamente 20% [Srivastava et al., 2014]. Essa diferença tão grande entre a taxa recomendada e a que resulta da aplicação de nossa técnica pode estar corrompendo demasiadamente os exemplos de treinamento e impedindo a criação de um modelo preditivo de melhor qualidade que tire proveito da alterações aleatórias sobre os exemplos de treinamento.

Em relação aos conjuntos do COIN (Teddy e Happy) não realizamos alterações para manter os resultados comparáveis ao trabalho de Beck et al. [2017b]. Os conjuntos de dados que coletamos diretamente do SDSS pelo portal *SciServer*, apresentaram valores atípicos, o que nos levou a aplicar diversos filtros (descritos na Seção 5.3.1). Dentre os exemplos removidos, estão os que apresentaram valores negativos em alguma das bandas ou erros (achamos valores de -9999 por exemplo), exemplos com erros maiores que 1.0 e exemplos com valor de magnitude na banda u maiores do que 25.

Ainda com relação aos conjuntos de dados e ao seu pré-processamento, adotamos para os modelos baseados em redes neurais o uso da normalização com o objetivo de melhorar seu poder preditivo. Aplicamos a normalização apenas às componentes de magnitudes, visto que as medidas de erro apresentaram valores pequenos. Embora os resultados não tenham sido reportados nesta dissertação, fizemos experimentos de treinamento de redes neurais sem aplicar normalização aos dados de entrada, e o efeito foi que os modelos resultantes foram significativamente piores. Por outro lado, para os modelos de regressão gerados por meio do algoritmo XGBoost, não aplicamos normalização por se tratar de um método baseado em árvores; nesse caso, constatamos que a aplicação ou não da normalização não influencia nos resultados.

Todo o código fonte desenvolvido nesta dissertação e os conjuntos de dados utilizados estão publicamente disponíveis. O código fonte que inclui a implementação de todos os métodos explorados neste trabalho se encontram no repositório `git` em <https://github.com/MLRG-CEFET-RJ/rna-cdropout>. Os conjuntos de dados do projeto COIN, tanto Teddy quanto Happy, estão disponíveis no repositório daquele projeto¹. O conjunto de dados do levantamento SDSS-DR16 pode ser obtido diretamente no portal do *SciServer*² após um breve cadastro³. De todo modo, também disponibilizamos no portal Zenodo os conjuntos de dados resultantes dos filtros que aplicamos durante nossa

¹https://github.com/COINtoolbox/photoz_catalogues

²<http://skyserver.sdss.org/dr16/en/tools/search/sql.aspx>

³<https://apps.sciserver.org/login-portal/register>

análise exploratória⁴.

6.2- Trabalhos Futuros

Na Seção 2.1.2, descrevemos o conceito de sistema de magnitude, fundamental para representar informações sobre objetos investigados na Astronomia, tais como galáxias. Nesta dissertação, utilizamos o sistema de magnitudes *ugriz* que foi adotado pelo SDSS. Uma possível via de continuação é realizar experimentos com conjunto de dados que usem outros sistemas de magnitudes, como por exemplo *UGRIZ* ou *u'g'r'i'z'*, diferentes do *ugriz* explorado neste trabalho.

Também pretendemos usar diferentes tamanhos de conjuntos de treinamento conforme criamos do SDSS-DR16 em SDSS-R16-120k. Usaríamos amostras de maiores tamanhos de 200 e 300 mil exemplos. Nosso objetivo com isso é investigar se a técnica de predição aqui proposta poderia se beneficiar de um conjunto de dados maior, uma vez que, por conta da natureza de nossa regularização, seriam criados ainda mais exemplos.

Conforme mencionado na Seção 6.1, constatamos nos experimentos que o uso de normalização dos dados foi fundamental para a geração de modelos de melhor qualidade no caso das redes neurais artificiais. Nesse sentido, planejamos testar as abordagens de redes neurais com outros normalizadores, como por exemplo o *MinMaxScaler* e o *MaxAbsScaler*. Provavelmente outros normalizadores tragam melhorias no poder preditivo dos modelos.

Está também em nossos planos adicionar outras *baselines* como de florestas aleatórias, *k*-NN, regressão polinomial, SVM e GLM. Desta forma poderíamos analisar potenciais métodos mais competitivos relativamente aos modelos gerados por redes MLP, obtendo assim uma conclusão mais fundamentada sobre o estado da arte da aprendizagem de máquina na tarefa de predizer foto-z.

Na técnica de regularização inspirada em *dropout* proposta nessa dissertação, é também possível explorarmos outras abordagens de alteração sobre os valores de magnitude. Por exemplo, no lugar de alterar para zero o valor de magnitude para as entradas menos confiáveis, poderíamos substituir por alguma estimativa da *magnitude*

⁴<https://zenodo.org/record/4291181>

esperada para o erro correspondente.

A técnica proposta nesta dissertação em essência produz perturbações aleatórias sobre os exemplos de treinamento passados à rede neural. Como efeito, a técnica realiza uma forma de aumento dos dados de treinamento (*data augmentation*). Nesse contexto, faz sentido realizar o aumento de dados antes do treinamento como uma atividade de pré-processamento. Assim seriam produzidos ainda mais exemplos para o treinamento e validação do modelo. Isto poderia melhorar o poder preditivo dos modelos.

Nesta dissertação, fixamos nossa investigação experimental sobre arquiteturas de redes neurais MLP com duas camadas ocultas. Testar modelos de redes neurais MLP com mais camadas ocultas também seria interessante, ampliando assim a possibilidade de gerar melhores modelos para prever foto-z.

Por fim, outra via de trabalho futuro é investigar a importância de considerar os erros de medição em outros problemas da Astronomia. Por exemplo, o problema conhecido como separação de estrelas e galáxias (*star-galaxy separation*) é um problema de classificação em aprendizado de máquina, no qual um modelo preditivo deve ser induzido para identificar a natureza de um objeto (estrela ou galáxia, nesse caso). Nesse contexto, o conjunto de dados disponibilizado pelo projeto *COSMOS*, no qual erros de medições também existem, podem ser usados [Soumagnac et al., 2015].

Apêndices

A- Resultados com outras Métricas

No Capítulo 5, apresentamos os resultados dos experimentos usando métricas de avaliação normalmente utilizadas na Astronomia em soluções para predição de *redshift* fotométrico que usam a abordagem de Aprendizado de Máquina. Essas métricas são descritas na Seção 5.4. Neste apêndice, apresentamos resultados dos mesmos modelos usados para produção dos resultados apresentados Capítulo 5. A diferença é que aqui apresentamos esses resultados usando métricas de avaliação normalmente utilizados em problemas de regressão em Aprendizado de Máquina.

Em particular, apresentamos os resultados usando as métricas seguintes: erro médio quadrático, raiz do erro médio quadrático e o erro médio absoluto [Sánchez et al., 2014; Way and Klose, 2012; Singh and Dayan, 1998]. Descrevemos essas medidas na Seção A.1 e apresentamos os resultados na Seção A.2.

A.1- MSE, MAE, RMSE

O erro médio quadrático (*Mean Squared Error* - MSE) é uma métrica que produz o valor ao quadrado associado a média dos erros ou perdas, usado como medida de qualidade de um estimador. Suas principais características são não produzir valores negativos e produzir valores próximos de zero [Lehmann and Casella, 2006]. Para o erro médio quadrático, seu valor é calculado conforme a Equação 22, onde m é a quantidade de amostras, \hat{y} e y são dois vetores m -dimensionais. As i -ésimas entradas em \hat{y} e y contêm o valor previsto e o valor efetivo para a i -ésima amostra, respectivamente.

$$\text{MSE}(\mathbf{y}, \hat{\mathbf{y}}) = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2 \quad (22)$$

O erro médio absoluto (*Mean Absolute Error* - MAE) é uma métrica que produz o valor da diferença entre duas variáveis, representando a distância horizontal média entre cada ponto e a linha de identidade, onde os valores de cada ponto representam o erro

ou a perda [Willmott and Matsuura, 2005]. Para o erro médio absoluto da função E , seu valor é calculado conforme na Equação 23, onde n é a quantidade de amostras, \hat{y} é o valor previsto da i -ésima amostra e y é o valor real relacionado.

$$\text{MAE}(\mathbf{y}, \hat{\mathbf{y}}) = \frac{1}{m} \sum_{i=1}^m |y_i - \hat{y}_i|. \quad (23)$$

A raiz do erro quadrático médio (*Root Mean Squared Error* - RMSE) corresponde à raiz quadrada do MSE, conforme a Equação 24.

$$\text{RMSE}(\mathbf{y}, \hat{\mathbf{y}}) = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2} \quad (24)$$

A.1.1- Coeficiente de determinação

O coeficiente de determinação, denotado por R^2 , corresponde à variância explicada pelo modelo de regressão (i.e., que pode ser corretamente prevista pelo modelo a partir das variáveis independentes). O valor igual a 1 corresponde à previsão perfeita. Se o modelo de regressão é ingênuo, i.e., se ele prediz \bar{y} para qualquer exemplo, então R^2 é igual a zero. O coeficiente também pode assumir valores negativos, quando o modelo de regressão tem qualidade pior do que o modelo de regressão ingênuo. O coeficiente de determinação é calculado usando a Equação 25.

$$R^2 \equiv 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}} \quad (25)$$

Na equação acima, SS_{tot} é soma total dos quadrados (*total sum of squares*). Esse valor é proporcional à variância de y . Se \hat{y} é a média dos componentes de y , então SS_{tot} é calculado conforme a seguir:

$$SS_{\text{tot}} = \sum_{i=1}^m [y_i - \bar{y}]^2$$

Já o valor SS_{res} corresponde à soma dos quadrados dos resíduos (*sum of squares of residuals*). SS_{res} é calculado conforme a seguir:

$$SS_{\text{res}} = \sum_{i=1}^m [y_i - \hat{y}_i]^2$$

A Figura 43 apresenta uma visão intuitiva sobre o significado do valor de R^2 . Quanto melhor o modelo de regressão (à direita) se ajustar os dados em comparação com a média simples (no gráfico à esquerda), mais próximo de 1 é o valor de R^2 . As áreas dos quadrados azuis (SS_{res}) representam os resíduos dos quadrados em relação à regressão linear. As áreas dos quadrados vermelhos (SS_{tot}) representam os resíduos dos quadrados em relação ao valor médio.

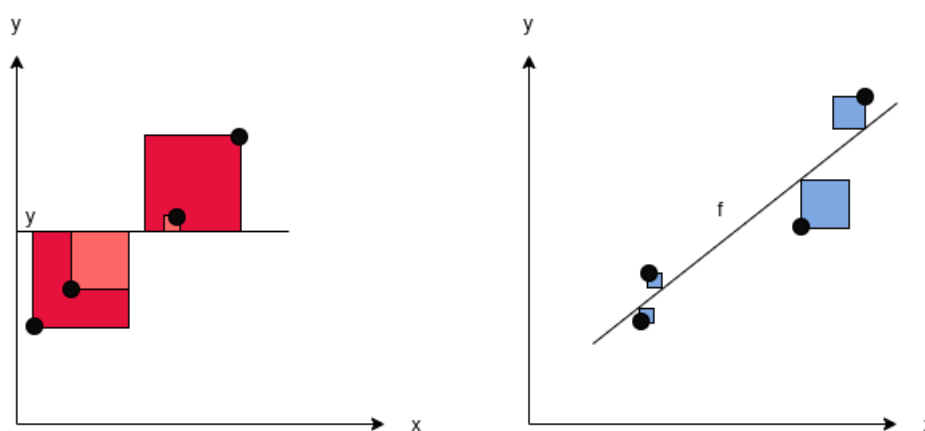


Figura 43 – Intuição sobre o significado do coeficiente R^2 . Adaptado de Orzetto¹.

A.2- Resultados

Nesta seção, apresentamos outra perspectiva dos resultados apresentados na Seção 5.6 nas Tabelas 12, 13 e 14 de forma complementar. Nestas incluímos os *scores* de MSE, RMSE, MAE e R2 conforme descritos na Seção A.1. A Tabela 18 mostra os resultados complementares dos modelos treinados no conjunto SDSS-R16-120k. A Tabela 19 mostra os resultados complementares dos modelos treinados para o conjunto COIN/Teddy. A Tabela 20 para os resultados por meio do COIN/Happy.

¹https://commons.wikimedia.org/wiki/File:Coefficient_of_Determination.svg

Método	MSE ($\times 10^{-2}$)	MAE ($\times 10^{-2}$)	RMSE ($\times 10^{-2}$)	MAD ($\times 10^{-2}$)	R2	Tempo (min)
RNA-RAD-Inv	1.35 \pm 0.04	5.41 \pm 0.24	11.61 \pm 0.17	1.87 \pm 0.16	0.8233 \pm 0.0053	67
RNA-RRI-Inv	1.34 \pm 0.04	5.39 \pm 0.26	11.58 \pm 0.18	1.88 \pm 0.17	0.8240 \pm 0.0055	65
RNA05	1.31 \pm 0.00	5.18 \pm 0.01	11.45 \pm 0.00	1.73 \pm 0.00	0.8281 \pm 0.0001	62
RNA10	1.26\pm0.01	5.04\pm0.05	11.22\pm0.04	1.70\pm0.05	0.8348\pm0.0011	64
XGB05	1.41 \pm 0.00	5.83 \pm 0.00	11.87 \pm 0.00	2.24 \pm 0.00	0.8151 \pm 0.0000	16
XGB10	1.35 \pm 0.00	5.73 \pm 0.00	11.63 \pm 0.00	2.22 \pm 0.00	0.8226 \pm 0.0001	31

Tabela 18 – Resultados complementares da Tabela 12, após 3.000 épocas, com 791 unidades na primeira camada oculta e 158 unidades na segunda camada oculta, usando o conjunto SDSS-R16-120k.

Método	MSE ($\times 10^{-2}$)	MAE ($\times 10^{-2}$)	RMSE ($\times 10^{-2}$)	MAD ($\times 10^{-2}$)	R2	Tempo (min)
B						
RNA-RAD-Inv	0.11±0.00	2.28±0.05	3.38±0.04	1.20±0.04	0.8515±0.0035	64
RNA-RRI-Inv	0.11±0.00	2.26±0.05	3.37±0.04	1.18±0.04	0.8529±0.0034	63
RNA05	0.12±0.01	2.37±0.12	3.48±0.11	1.26±0.09	0.8427±0.0104	62
RNA10	0.11±0.00	2.18±0.01	3.29±0.03	1.13±0.01	0.8598±0.0024	60
XGB05	0.12±0.00	2.40±0.00	3.48±0.00	1.31±0.00	0.8428±0.0000	12
XGB10	0.12±0.00	2.39±0.00	3.46±0.00	1.31±0.00	0.8451±0.0000	24
C						
RNA-RAD-Inv	0.13±0.00	2.38±0.07	3.63±0.05	1.20±0.05	0.8670±0.0034	64
RNA-RRI-Inv	0.13±0.00	2.36±0.07	3.61±0.05	1.18±0.06	0.8684±0.0035	63
RNA05	0.14±0.01	2.49±0.15	3.74±0.14	1.28±0.12	0.8587±0.0103	62
RNA10	0.12±0.00	2.27±0.01	3.51±0.01	1.13±0.01	0.8756±0.0004	60
XGB05	0.14±0.00	2.55±0.00	3.77±0.00	1.36±0.00	0.8570±0.0000	12
XGB10	0.14±0.00	2.55±0.00	3.74±0.00	1.37±0.00	0.8586±0.0000	24
D						
RNA-RAD-Inv	2.05±2.18	6.99±3.56	12.89±6.26	2.76±1.06	0.5656±0.4615	64
RNA-RRI-Inv	2.49±2.22	7.43±3.39	14.44±6.32	2.81±1.02	0.4744±0.4688	63
RNA05	1.10±0.70	5.98±0.86	10.13±2.78	2.88±0.51	0.7668±0.1475	62
RNA10	3.28±2.99	7.78±3.07	16.74±6.89	2.49±0.26	0.3069±0.6321	60
XGB05	1.18±0.00	8.07±0.00	10.84±0.00	5.02±0.00	0.7515±0.0001	12
XGB10	1.00±0.00	7.65±0.01	9.98±0.01	4.91±0.00	0.7893±0.0003	24

Tabela 19 – Resultados complementares da Tabela 13, após 3.000 épocas, com 791 unidades na primeira camada oculta e 158 unidades na segunda camada oculta, usando o conjunto COIN/Teddy A para treinamento/validação e B,C e D para teste.

Método	MSE ($\times 10^{-2}$)	MAE ($\times 10^{-2}$)	RMSE ($\times 10^{-2}$)	MAD ($\times 10^{-2}$)	R2	Tempo (min)
B						
RNA-RAD-Inv	0.31±0.02	3.24±0.14	5.57±0.15	1.57±0.11	0.9304±0.0038	66
RNA-RRI-Inv	0.30±0.02	3.17±0.13	5.51±0.14	1.52±0.10	0.9319±0.0034	65
RNA05	0.29±0.00	3.08±0.01	5.42±0.03	1.46±0.02	0.9343±0.0007	62
RNA10	0.28±0.01	3.01±0.02	5.33±0.10	1.42±0.02	0.9363±0.0024	62
XGB05	0.33±0.00	3.49±0.00	5.70±0.00	1.77±0.00	0.9272±0.0000	12
XGB10	0.32±0.00	3.52±0.00	5.67±0.00	1.81±0.00	0.9281±0.0000	24
C						
RNA-RAD-Inv	1.24±0.03	7.42±0.17	11.14±0.15	3.71±0.15	0.6291±0.0099	66
RNA-RRI-Inv	1.23±0.02	7.32±0.14	11.07±0.10	3.63±0.13	0.6334±0.0069	65
RNA05	1.21±0.00	7.20±0.00	11.00±0.01	3.53±0.00	0.6381±0.0009	62
RNA10	1.13±0.01	6.91±0.06	10.63±0.06	3.37±0.05	0.6623±0.0037	62
XGB05	1.40±0.00	8.20±0.00	11.85±0.00	4.19±0.00	0.5803±0.0000	12
XGB10	1.36±0.00	8.02±0.00	11.68±0.00	4.05±0.00	0.5923±0.0000	24
D						
RNA-RAD-Inv	3.14±0.09	11.92±0.16	17.72±0.26	5.64±0.12	0.3556±0.0186	66
RNA-RRI-Inv	3.18±0.10	11.93±0.14	17.82±0.28	5.59±0.09	0.3477±0.0209	65
RNA05	3.15±0.03	11.85±0.10	17.75±0.07	5.54±0.07	0.3532±0.0053	62
RNA10	4.76±2.58	11.45±0.31	21.19±5.19	5.21±0.11	0.0226±0.5305	62
XGB05	3.12±0.00	12.30±0.00	17.65±0.00	5.94±0.00	0.3605±0.0001	12
XGB10	3.04±0.00	12.15±0.00	17.45±0.00	5.82±0.00	0.3750±0.0000	24

Tabela 20 – Resultados complementares da Tabela 14, após 3.000 épocas, com 791 unidades na primeira camada oculta e 158 unidades na segunda camada oculta, usando o conjunto COIN/Happy A para treinamento/validação e B, C e D para teste.

B- Resultados com outras Quantidades de Épocas

Neste apêndice, apresentamos resultados usando as métricas introduzidas na Seção A.1. Assim como no Apêndice A, os resultados das predições são apresentados com as métricas MSE, RMSE, MAE e R2. Desta vez, variamos as quantidades de épocas de treinamento conforme sumário abaixo. Os tempos médios de treinamento também são fornecidos (em minutos).

- 500 épocas: Tabela 23, Tabela 22 e Tabela 21;
- 1000 épocas: Tabela 26, Tabela 25 e Tabela 24;
- 2000 épocas: Tabela 29, Tabela 28 e Tabela 27.

Método	MSE ($\times 10^{-2}$)	MAE ($\times 10^{-2}$)	RMSE ($\times 10^{-2}$)	MAD ($\times 10^{-2}$)	R2	Tempo (min)
RNA-RAD-Inv	1.40 \pm 0.01	5.81 \pm 0.09	11.83 \pm 0.03	2.17 \pm 0.10	0.8165 \pm 0.0010	12
RNA-RRI-Inv	1.40 \pm 0.01	5.83 \pm 0.09	11.85 \pm 0.04	2.19 \pm 0.12	0.8158 \pm 0.0012	12
RNA05	1.31 \pm 0.00	5.18 \pm 0.01	11.45 \pm 0.00	1.73 \pm 0.00	0.8281 \pm 0.0001	10
RNA10	1.26\pm0.01	5.04\pm0.05	11.22\pm0.04	1.70\pm0.05	0.8348\pm0.0011	11
XGB05	3.73 \pm 0.00	15.40 \pm 0.00	19.31 \pm 0.00	10.60 \pm 0.00	0.5111 \pm 0.0000	3
XGB10	3.68 \pm 0.00	15.32 \pm 0.00	19.17 \pm 0.00	10.57 \pm 0.00	0.5178 \pm 0.0000	5

Tabela 21 – Resultados após 500 épocas, com 791 unidades na primeira camada oculta e 158 unidades na segunda camada oculta, usando o conjunto SDSS-R16-120k.

Método	MSE ($\times 10^{-2}$)	MAE ($\times 10^{-2}$)	RMSE ($\times 10^{-2}$)	MAD ($\times 10^{-2}$)	R2	Tempo (min)
B						
RNA-RAD-Inv	0.12±0.00	2.36±0.02	3.44±0.01	1.27±0.03	0.8468±0.0006	11
RNA-RRI-Inv	0.12±0.00	2.35±0.02	3.44±0.01	1.26±0.03	0.8469±0.0008	11
RNA05	0.11±0.00	2.23±0.01	3.34±0.00	1.16±0.01	0.8554±0.0001	10
RNA10	0.11±0.00	2.18±0.01	3.28±0.03	1.13±0.01	0.8601±0.0024	10
XGB05	0.37±0.00	4.89±0.00	6.06±0.00	3.34±0.00	0.5235±0.0000	2
XGB10	0.37±0.00	4.90±0.00	6.08±0.00	3.33±0.00	0.5209±0.0000	4
C						
RNA-RAD-Inv	0.14±0.00	2.48±0.03	3.69±0.01	1.29±0.04	0.8625±0.0008	11
RNA-RRI-Inv	0.14±0.00	2.47±0.03	3.69±0.01	1.28±0.05	0.8627±0.0010	11
RNA05	0.13±0.00	2.32±0.01	3.58±0.00	1.15±0.01	0.8709±0.0003	10
RNA10	0.12±0.00	2.27±0.01	3.51±0.01	1.13±0.01	0.8757±0.0004	10
XGB05	0.47±0.00	5.69±0.00	6.86±0.00	4.13±0.00	0.5251±0.0000	2
XGB10	0.47±0.00	5.70±0.00	6.88±0.00	4.15±0.00	0.5219±0.0000	4
D						
RNA-RAD-Inv	2.15±2.18	6.99±3.56	12.89±6.26	2.79±1.08	0.5542±0.4631	11
RNA-RRI-Inv	2.51±2.24	7.46±3.42	14.48±6.33	2.88±1.02	0.4643±0.4697	11
RNA05	1.04±0.13	5.07±0.12	10.19±0.64	2.07±0.05	0.7793±0.0278	10
RNA10	2.39±1.40	6.97±1.84	14.93±4.05	2.44±0.24	0.4939±0.2966	10
XGB05	2.76±0.00	14.60±0.00	16.62±0.00	11.07±0.00	0.4157±0.0001	2
XGB10	2.67±0.00	14.47±0.00	16.33±0.00	11.01±0.00	0.4361±0.0003	4

Tabela 22 – Resultados após 500 épocas com 791 unidades na primeira camada oculta e 158 unidades na segunda camada oculta, usando o conjunto COIN/Teddy A para treinamento/validação e B, C e D para teste.

Método	MSE ($\times 10^{-2}$)	MAE ($\times 10^{-2}$)	RMSE ($\times 10^{-2}$)	MAD ($\times 10^{-2}$)	R2	Tempo (min)
B						
RNA-RAD-Inv	0.33±0.00	3.46±0.04	5.76±0.03	1.73±0.04	0.9258±0.0007	12
RNA-RRI-Inv	0.33±0.00	3.47±0.06	5.78±0.04	1.74±0.05	0.9253±0.0011	11
RNA05	0.29±0.00	3.08±0.01	5.42±0.03	1.46±0.02	0.9343±0.0007	10
RNA10	0.28±0.01	3.01±0.02	5.33±0.10	1.42±0.02	0.9363±0.0024	10
XGB05	1.88±0.00	11.91±0.00	13.73±0.00	8.70±0.00	0.5781±0.0000	2
XGB10	1.89±0.00	11.93±0.00	13.76±0.00	8.73±0.00	0.5759±0.0000	4
C						
RNA-RAD-Inv	1.27±0.03	7.65±0.11	11.27±0.11	3.93±0.09	0.6201±0.0076	12
RNA-RRI-Inv	1.26±0.01	7.62±0.06	11.23±0.05	3.93±0.09	0.6230±0.0031	11
RNA05	1.21±0.00	7.20±0.00	11.00±0.01	3.53±0.00	0.6381±0.0009	10
RNA10	1.13±0.01	6.91±0.06	10.63±0.06	3.37±0.05	0.6623±0.0037	10
XGB05	2.09±0.00	11.26±0.00	14.45±0.00	6.99±0.00	0.3759±0.0000	2
XGB10	2.10±0.00	11.26±0.00	14.48±0.00	6.92±0.00	0.3731±0.0000	4
D						
RNA-RAD-Inv	3.17±0.07	12.02±0.13	17.81±0.19	5.78±0.10	0.3486±0.0138	12
RNA-RRI-Inv	3.15±0.08	11.94±0.12	17.76±0.22	5.69±0.08	0.3527±0.0160	11
RNA05	3.15±0.03	11.85±0.10	17.75±0.07	5.54±0.07	0.3532±0.0053	10
RNA10	4.76±2.58	11.45±0.31	21.19±5.19	5.21±0.11	0.0226±0.5305	10
XGB05	4.09±0.00	15.20±0.00	20.23±0.00	8.82±0.00	0.1600±0.0001	2
XGB10	3.94±0.00	14.85±0.00	19.84±0.00	8.47±0.00	0.1920±0.0000	4

Tabela 23 – Resultados após 500 épocas, com 791 unidades na primeira camada oculta e 158 unidades na segunda camada oculta, usando o conjunto COIN/Happy A para treinamento/validação e B, C e D para teste.

Método	MSE ($\times 10^{-2}$)	MAE ($\times 10^{-2}$)	RMSE ($\times 10^{-2}$)	MAD ($\times 10^{-2}$)	R2	Tempo (min)
RNA-RAD-Inv	1.40 \pm 0.01	5.77 \pm 0.06	11.82 \pm 0.03	2.14 \pm 0.09	0.8167 \pm 0.0010	24
RNA-RR1-Inv	1.40 \pm 0.01	5.82 \pm 0.09	11.84 \pm 0.04	2.18 \pm 0.12	0.8161 \pm 0.0013	24
RNA05	1.31 \pm 0.00	5.18 \pm 0.01	11.45 \pm 0.00	1.73 \pm 0.00	0.8281 \pm 0.0001	21
RNA10	1.26\pm0.01	5.04\pm0.05	11.22\pm0.04	1.70\pm0.05	0.8348\pm0.0011	21
XGB05	2.28 \pm 0.00	10.72 \pm 0.00	15.10 \pm 0.00	6.68 \pm 0.00	0.7010 \pm 0.0000	5
XGB10	2.22 \pm 0.00	10.61 \pm 0.00	14.89 \pm 0.00	6.66 \pm 0.00	0.7092 \pm 0.0000	10

Tabela 24 – Resultados após 1.000 épocas, com 791 unidades na primeira camada oculta e 158 unidades na segunda camada oculta, usando o conjunto SDSS-R16-120k.

Método	MSE ($\times 10^{-2}$)	MAE ($\times 10^{-2}$)	RMSE ($\times 10^{-2}$)	MAD ($\times 10^{-2}$)	R2	Tempo (min)
B						
RNA-RAD-Inv	0.12±0.00	2.36±0.02	3.43±0.01	1.27±0.03	0.8474±0.0009	23
RNA-RRI-Inv	0.12±0.00	2.36±0.02	3.43±0.01	1.28±0.03	0.8470±0.0006	23
RNA05	0.11±0.00	2.23±0.01	3.34±0.00	1.16±0.01	0.8550±0.0003	20
RNA10	0.11±0.00	2.18±0.01	3.29±0.03	1.13±0.01	0.8598±0.0024	20
XGB05	0.22±0.00	3.63±0.00	4.64±0.00	2.37±0.00	0.7206±0.0000	4
XGB10	0.22±0.00	3.64±0.00	4.66±0.00	2.36±0.00	0.7183±0.0000	8
C						
RNA-RAD-Inv	0.14±0.00	2.48±0.03	3.69±0.01	1.30±0.04	0.8628±0.0010	23
RNA-RRI-Inv	0.14±0.00	2.49±0.03	3.70±0.01	1.30±0.05	0.8622±0.0010	23
RNA05	0.13±0.00	2.32±0.01	3.58±0.00	1.15±0.01	0.8707±0.0000	20
RNA10	0.12±0.00	2.27±0.01	3.51±0.01	1.13±0.01	0.8756±0.0004	20
XGB05	0.27±0.00	4.14±0.00	5.20±0.00	2.82±0.00	0.7271±0.0000	4
XGB10	0.27±0.00	4.17±0.00	5.23±0.00	2.84±0.00	0.7245±0.0000	8
D						
RNA-RAD-Inv	2.15±2.18	6.99±3.56	12.89±6.26	2.79±1.08	0.5542±0.4631	23
RNA-RRI-Inv	2.51±2.24	7.46±3.42	14.48±6.33	2.88±1.02	0.4643±0.4697	23
RNA05	2.03±0.85	6.30±1.11	13.91±3.07	2.26±0.14	0.5711±0.1804	20
RNA10	3.28±2.99	7.78±3.07	16.74±6.89	2.49±0.26	0.3069±0.6321	20
XGB05	1.91±0.00	11.57±0.00	13.82±0.00	8.34±0.00	0.5959±0.0002	4
XGB10	1.77±0.00	11.36±0.00	13.32±0.00	8.32±0.00	0.6250±0.0002	8

Tabela 25 – Resultados após 1.000 épocas, com 791 unidades na primeira camada oculta e 158 unidades na segunda camada oculta, usando o conjunto COIN/Teddy A para treinamento/validação e B, C e D para teste.

Método	MSE ($\times 10^{-2}$)	MAE ($\times 10^{-2}$)	RMSE ($\times 10^{-2}$)	MAD ($\times 10^{-2}$)	R2	Tempo (min)
B						
RNA-RAD-Inv	0.33±0.01	3.39±0.04	5.74±0.06	1.69±0.05	0.9262±0.0015	22
RNA-RRI-Inv	0.33±0.01	3.39±0.06	5.77±0.07	1.69±0.07	0.9255±0.0017	22
RNA05	0.29±0.00	3.08±0.01	5.42±0.03	1.46±0.02	0.9343±0.0007	21
RNA10	0.28±0.01	3.01±0.02	5.33±0.10	1.42±0.02	0.9363±0.0024	20
XGB05	0.91±0.00	7.80±0.00	9.56±0.00	5.31±0.00	0.7952±0.0000	4
XGB10	0.93±0.00	7.86±0.00	9.62±0.00	5.39±0.00	0.7926±0.0000	8
C						
RNA-RAD-Inv	1.28±0.02	7.60±0.08	11.32±0.09	3.85±0.07	0.6169±0.0060	22
RNA-RRI-Inv	1.28±0.02	7.58±0.09	11.30±0.09	3.84±0.07	0.6184±0.0062	22
RNA05	1.21±0.00	7.20±0.00	11.00±0.01	3.53±0.00	0.6381±0.0009	21
RNA10	1.13±0.01	6.91±0.06	10.63±0.06	3.37±0.05	0.6623±0.0037	20
XGB05	1.65±0.00	9.56±0.00	12.85±0.00	5.52±0.00	0.5062±0.0000	4
XGB10	1.67±0.00	9.58±0.00	12.91±0.00	5.51±0.00	0.5022±0.0000	8
D						
RNA-RAD-Inv	3.17±0.07	12.02±0.13	17.81±0.19	5.78±0.10	0.3486±0.0138	22
RNA-RRI-Inv	3.15±0.08	11.94±0.12	17.76±0.22	5.69±0.08	0.3527±0.0160	22
RNA05	3.15±0.03	11.85±0.10	17.75±0.07	5.54±0.07	0.3532±0.0053	21
RNA10	4.76±2.58	11.45±0.31	21.19±5.19	5.21±0.11	0.0226±0.5305	20
XGB05	3.41±0.00	13.43±0.00	18.48±0.00	7.24±0.00	0.2993±0.0001	4
XGB10	3.32±0.00	13.25±0.00	18.23±0.00	7.07±0.00	0.3180±0.0000	8

Tabela 26 – Resultados após 1.000 épocas, com 791 unidades na primeira camada oculta e 158 unidades na segunda camada oculta, usando o conjunto COIN/Happy A para treinamento/validação e B, C e D para teste.

Método	MSE ($\times 10^{-2}$)	MAE ($\times 10^{-2}$)	RMSE ($\times 10^{-2}$)	MAD ($\times 10^{-2}$)	R2	Tempo (min)
RNA-RAD-Inv	1.40 \pm 0.01	5.77 \pm 0.06	11.82 \pm 0.03	2.14 \pm 0.09	0.8167 \pm 0.0010	48
RNA-RRI-Inv	1.40 \pm 0.01	5.82 \pm 0.09	11.84 \pm 0.04	2.18 \pm 0.12	0.8161 \pm 0.0013	47
RNA05	1.31 \pm 0.00	5.18 \pm 0.01	11.45 \pm 0.00	1.73 \pm 0.00	0.8281 \pm 0.0001	42
RNA10	1.26\pm0.01	5.04\pm0.05	11.22\pm0.04	1.70\pm0.05	0.8348\pm0.0011	42
XGB05	1.52 \pm 0.00	6.77 \pm 0.00	12.34 \pm 0.00	3.09 \pm 0.00	0.8001 \pm 0.0000	10
XGB10	1.47 \pm 0.00	6.66 \pm 0.00	12.11 \pm 0.00	3.08 \pm 0.00	0.8076 \pm 0.0000	20

Tabela 27 – Resultados após 2.000 épocas, com 791 unidades na primeira camada oculta e 158 unidades na segunda camada oculta, usando o conjunto SDSS-R16-120k.

Método	MSE ($\times 10^{-2}$)	MAE ($\times 10^{-2}$)	RMSE ($\times 10^{-2}$)	MAD ($\times 10^{-2}$)	R2	Tempo (min)
B						
RNA-RAD-Inv	0.12±0.00	2.36±0.02	3.43±0.01	1.28±0.03	0.8474±0.0008	45
RNA-RRI-Inv	0.12±0.00	2.36±0.02	3.43±0.01	1.28±0.03	0.8470±0.0006	45
RNA05	0.11±0.00	2.23±0.01	3.34±0.00	1.16±0.01	0.8550±0.0003	40
RNA10	0.11±0.00	2.18±0.01	3.28±0.02	1.13±0.01	0.8604±0.0020	40
XGB05	0.13±0.00	2.64±0.00	3.66±0.00	1.54±0.00	0.8259±0.0000	8
XGB10	0.13±0.00	2.64±0.00	3.66±0.00	1.54±0.00	0.8264±0.0000	16
C						
RNA-RAD-Inv	0.14±0.00	2.48±0.03	3.69±0.01	1.30±0.04	0.8628±0.0009	45
RNA-RRI-Inv	0.14±0.00	2.49±0.03	3.69±0.01	1.31±0.05	0.8623±0.0010	45
RNA05	0.13±0.00	2.32±0.01	3.58±0.00	1.15±0.01	0.8707±0.0000	40
RNA10	0.12±0.00	2.27±0.01	3.51±0.01	1.13±0.01	0.8756±0.0005	40
XGB05	0.16±0.00	2.87±0.00	4.01±0.00	1.68±0.00	0.8379±0.0000	8
XGB10	0.16±0.00	2.89±0.00	4.01±0.00	1.69±0.00	0.8380±0.0000	16
D						
RNA-RAD-Inv	2.15±2.19	6.99±3.56	12.89±6.26	2.79±1.08	0.5556±0.4635	45
RNA-RRI-Inv	2.50±2.22	7.46±3.42	14.47±6.33	2.88±1.02	0.4644±0.4699	45
RNA05	2.03±0.85	6.30±1.11	13.91±3.07	2.26±0.14	0.5711±0.1804	40
RNA10	1.47±0.55	5.87±0.87	11.91±2.19	2.32±0.28	0.6900±0.1157	40
XGB05	1.34±0.00	8.99±0.00	11.60±0.00	5.94±0.00	0.7157±0.0001	8
XGB10	1.16±0.00	8.57±0.00	10.76±0.01	5.80±0.00	0.7550±0.0002	16

Tabela 28 – Resultados após 2.000 épocas, com 791 unidades na primeira camada oculta e 158 unidades na segunda camada oculta, usando o conjunto COIN/Teddy A para treinamento/validação e B,C e D para teste.

Método	MSE ($\times 10^{-2}$)	MAE ($\times 10^{-2}$)	RMSE ($\times 10^{-2}$)	MAD ($\times 10^{-2}$)	R2	Tempo (min)
B						
RNA-RAD-Inv	0.33±0.01	3.38±0.04	5.73±0.06	1.68±0.05	0.9264±0.0016	46
RNA-RRI-Inv	0.33±0.01	3.38±0.06	5.77±0.07	1.68±0.07	0.9254±0.0017	46
RNA05	0.29±0.00	3.08±0.01	5.42±0.03	1.46±0.02	0.9343±0.0007	41
RNA10	0.28±0.01	3.01±0.02	5.33±0.10	1.42±0.02	0.9363±0.0024	41
XGB05	0.41±0.00	4.32±0.00	6.37±0.00	2.44±0.00	0.9093±0.0000	8
XGB10	0.41±0.00	4.40±0.00	6.39±0.00	2.54±0.00	0.9084±0.0000	16
C						
RNA-RAD-Inv	1.28±0.02	7.57±0.06	11.29±0.08	3.84±0.05	0.6188±0.0056	46
RNA-RRI-Inv	1.27±0.02	7.55±0.06	11.28±0.08	3.83±0.04	0.6197±0.0054	46
RNA05	1.21±0.00	7.20±0.00	11.00±0.01	3.53±0.00	0.6381±0.0009	41
RNA10	1.13±0.01	6.91±0.06	10.63±0.06	3.37±0.05	0.6623±0.0037	41
XGB05	1.46±0.00	8.52±0.00	12.06±0.00	4.50±0.00	0.5651±0.0000	8
XGB10	1.44±0.00	8.47±0.00	12.01±0.00	4.44±0.00	0.5687±0.0000	16
D						
RNA-RAD-Inv	3.16±0.07	11.99±0.14	17.77±0.21	5.75±0.09	0.3516±0.0150	46
RNA-RRI-Inv	3.15±0.08	11.93±0.10	17.76±0.22	5.67±0.04	0.3528±0.0159	46
RNA05	3.15±0.03	11.85±0.10	17.75±0.07	5.54±0.07	0.3532±0.0053	41
RNA10	4.76±2.58	11.45±0.31	21.19±5.19	5.21±0.11	0.0226±0.5305	41
XGB05	3.13±0.00	12.47±0.00	17.69±0.00	6.21±0.00	0.3577±0.0001	8
XGB10	3.08±0.00	12.36±0.00	17.54±0.00	6.10±0.00	0.3685±0.0000	16

Tabela 29 – Resultados após 2.000 épocas, com 791 unidades na primeira camada oculta e 158 unidades na segunda camada oculta, usando o conjunto COIN/Happy A para treinamento/validação e B,C e D para teste.

Referências Bibliográficas

- Bachega, R. R. A. (2014). *Vinculando modelos de energia escura com idade de galáxias em altos redshifts*. PhD thesis, Universidade de São Paulo. (Citações: 28.)
- Bahcall, N. A. (2015). Hubble's law and the expanding universe. *Proceedings of the National Academy of Sciences*, 112(11):3173–3175. (Citações: 26.)
- Ball, N. M. and Brunner, R. J. (2010). Data mining and machine learning in astronomy. *International Journal of Modern Physics D*, 19(07):1049–1106. (Citações: 14.)
- Barlow, R. E. (1972). Statistical inference under order restrictions; the theory and application of isotonic regression. Technical report. (Citações: 39.)
- Beck, R., Dobos, L., Budavári, T., Szalay, A. S., and Csabai, I. (2017a). Photo-z-sql: Integrated, flexible photometric redshift computation in a database. *Astronomy and computing*, 19:34–44. (Citações: 31 e 81.)
- Beck, R., Lin, C.-A., Ishida, E., Gieseke, F., de Souza, R., Costa-Duarte, M., Hattab, M., Krone-Martins, A., and Collaboration, C. (2017b). On the realistic validation of photometric redshifts. *Monthly Notices of the Royal Astronomical Society*, 468(4):4323–4339. (Citações: 43, 45, 46, 57, 72, 76, 78, 79, 90, 92 e 100.)
- Bezerra, E. (2016). Introdução à aprendizagem profunda. In Ogasawara, E. and Ferreira, V., editors, *Tópicos em Gerenciamento de Dados e Informações.*, chapter 3, pages 57–86. Editora SBC. (Citações: 34.)
- Bilicki, M., Hoekstra, H., Amaro, V., Blake, C., Brown, M., Cavuoti, S., de Jong, J., Hildebrandt, H., Wolf, C., Amon, A., et al. (2017). Photometric redshifts for the kilo-degree survey. machine-learning analysis with artificial neural networks. *arXiv preprint arXiv:1709.04205*. (Citações: 16.)
- Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, page 785–794, New York, NY, USA. Association for Computing Machinery. (Citações: 18, 37 e 38.)

Comins, N. F. and Kaufmann III, W. J. (2010). *Descobrimos o universo*. (Citações: 15, 21, 28 e 29.)

Dawson, K. S., Schlegel, D. J., Ahn, C. P., Anderson, S. F., Aubourg, É., Bailey, S., Barkhouser, R. H., Bautista, J. E., Beifiori, A., Berlind, A. A., Bhardwaj, V., Bizyaev, D., Blake, C. H., Blanton, M. R., Blomqvist, M., Bolton, A. S., Borde, A., Bovy, J., Brandt, W. N., Brewington, H., Brinkmann, J., Brown, P. J., Brownstein, J. R., Bundy, K., Busca, N. G., Carithers, W., Carnero, A. R., Carr, M. A., Chen, Y., Comparat, J., Connolly, N., Cope, F., Croft, R. A. C., Cuesta, A. J., da Costa, L. N., Davenport, J. R. A., Delubac, T., de Putter, R., Dhital, S., Ealet, A., Ebelke, G. L., Eisenstein, D. J., Escoffier, S., Fan, X., Filiz Ak, N., Finley, H., Font-Ribera, A., Génova-Santos, R., Gunn, J. E., Guo, H., Haggard, D., Hall, P. B., Hamilton, J.-C., Harris, B., Harris, D. W., Ho, S., Hogg, D. W., Holder, D., Honscheid, K., Huehnerhoff, J., Jordan, B., Jordan, W. P., Kauffmann, G., Kazin, E. A., Kirkby, D., Klaene, M. A., Kneib, J.-P., Le Goff, J.-M., Lee, K.-G., Long, D. C., Loomis, C. P., Lundgren, B., Lupton, R. H., Maia, M. A. G., Makler, M., Malanushenko, E., Malanushenko, V., Mandelbaum, R., Manera, M., Maraston, C., Margala, D., Masters, K. L., McBride, C. K., McDonald, P., McGreer, I. D., McMahon, R. G., Mena, O., Miralda-Escudé, J., Montero-Dorta, A. D., Montesano, F., Muna, D., Myers, A. D., Naugle, T., Nichol, R. C., Noterdaeme, P., Nuza, S. E., Olmstead, M. D., Oravetz, A., Oravetz, D. J., Owen, R., Padmanabhan, N., Palanque-Delabrouille, N., Pan, K., Parejko, J. K., Pâris, I., Percival, W. J., Pérez-Fournon, I., Pérez-Ràfols, I., Petitjean, P., Pfaffenberger, R., Pforr, J., Pieri, M. M., Prada, F., Price-Whelan, A. M., Raddick, M. J., Rebolo, R., Rich, J., Richards, G. T., Rockosi, C. M., Roe, N. A., Ross, A. J., Ross, N. P., Rossi, G., Rubiño-Martin, J. A., Samushia, L., Sánchez, A. G., Sayres, C., Schmidt, S. J., Schneider, D. P., Scóccola, C. G., Seo, H.-J., Shelden, A., Sheldon, E., Shen, Y., Shu, Y., Slosar, A., Smee, S. A., Snedden, S. A., Stauffer, F., Steele, O., Strauss, M. A., Streblyanska, A., Suzuki, N., Swanson, M. E. C., Tal, T., Tanaka, M., Thomas, D., Tinker, J. L., Tojeiro, R., Tremonti, C. A., Vargas Magaña, M., Verde, L., Viel, M., Wake, D. A., Watson, M., Weaver, B. A., Weinberg, D. H., Weiner, B. J., West, A. A., White, M., Wood-Vasey, W. M., Yeche, C., Zehavi, I., Zhao, G.-B., and Zheng, Z. (2013). *The Baryon Oscillation Spectroscopic Survey of SDSS-III*. , 145(1):10. (Citações: 31.)

De Wei, K. C. and Yang, A. (2019). Photometric redshift analysis using supervised learning algorithms and deep learning. In *EPJ Web of Conferences*, volume 206, page 09006.

EDP Sciences. (Citações: 43, 45 e 46.)

DESI Collaboration, Aghamousa, A., Aguilar, J., Ahlen, S., Alam, S., Allen, L. E., Allende Prieto, C., Annis, J., Bailey, S., Balland, C., Ballester, O., Baltay, C., Beaufore, L., Bebek, C., Beers, T. C., Bell, E. F., Bernal, J. L., Besuner, R., Beutler, F., Blake, C., Bleuler, H., Blomqvist, M., Blum, R., Bolton, A. S., Briceno, C., Brooks, D., Brownstein, J. R., Buckley-Geer, E., Burden, A., Burtin, E., Busca, N. G., Cahn, R. N., Cai, Y.-C., Cardiel-Sas, L., Carlberg, R. G., Carton, P.-H., Casas, R., Castander, F. J., Cervantes-Cota, J. L., Claybaugh, T. M., Close, M., Coker, C. T., Cole, S., Comparat, J., Cooper, A. P., Cousinou, M. C., Crocce, M., Cuby, J.-G., Cunningham, D. P., Davis, T. M., Dawson, K. S., de la Macorra, A., De Vicente, J., Delubac, T., Derwent, M., Dey, A., Dhungana, G., Ding, Z., Doel, P., Duan, Y. T., Ealet, A., Edelstein, J., Eftekhazadeh, S., Eisenstein, D. J., Elliott, A., Escoffier, S., Evatt, M., Fagrelus, P., Fan, X., Fanning, K., Farahi, A., Farihi, J., Favole, G., Feng, Y., Fernandez, E., Findlay, J. R., Finkbeiner, D. P., Fitzpatrick, M. J., Flaughner, B., Flender, S., Font-Ribera, A., Forero-Romero, J. E., Fosalba, P., Frenk, C. S., Fumagalli, M., Gaensicke, B. T., Gallo, G., Garcia-Bellido, J., Gaztanaga, E., Pietro Gentile Fusillo, N., Gerard, T., Gershkovich, I., Giannantonio, T., Gillet, D., Gonzalez-de-Rivera, G., Gonzalez-Perez, V., Gott, S., Graur, O., Gutierrez, G., Guy, J., Habib, S., Heetderks, H., Heetderks, I., Heitmann, K., Hellwing, W. A., Herrera, D. A., Ho, S., Holland, S., Honscheid, K., Huff, E., Hutchinson, T. A., Huterer, D., Hwang, H. S., Illa Laguna, J. M., Ishikawa, Y., Jacobs, D., Jeffrey, N., Jelinsky, P., Jennings, E., Jiang, L., Jimenez, J., Johnson, J., Joyce, R., Jullo, E., Juneau, S., Kama, S., Karcher, A., Karkar, S., Kehoe, R., Kennamer, N., Kent, S., Kilbinger, M., Kim, A. G., Kirkby, D., Kisner, T., Kitanidis, E., Kneib, J.-P., Kuposov, S., Kovacs, E., Koyama, K., Kremin, A., Kron, R., Kronig, L., Kueter-Young, A., Lacey, C. G., Lafever, R., Lahav, O., Lambert, A., Lampton, M., Landriau, M., Lang, D., Lauer, T. R., Le Goff, J.-M., Le Guillou, L., Le Van Suu, A., Lee, J. H., Lee, S.-J., Leitner, D., Lesser, M., Levi, M. E., L'Huillier, B., Li, B., Liang, M., Lin, H., Linder, E., Loebman, S. R., Lukić, Z., Ma, J., MacCrann, N., Magneville, C., Makarem, L., Manera, M., Manser, C. J., Marshall, R., Martini, P., Massey, R., Matheson, T., McCauley, J., McDonald, P., McGreer, I. D., Meisner, A., Metcalfe, N., Miller, T. N., Miquel, R., Moustakas, J., Myers, A., Naik, M., Newman, J. A., Nichol, R. C., Nicola, A., Nicolati da Costa, L., Nie, J., Niz, G., Norberg, P., Nord, B., Norman, D., Nugent, P., O'Brien, T., Oh, M., Olsen, K. A. G., Padilla, C., Padmanabhan, H., Padmanabhan, N., Palanque-Delabrouille, N., Palmese, A., Pappalardo, D., Pâris,

I., Park, C., Patej, A., Peacock, J. A., Peiris, H. V., Peng, X., Percival, W. J., Perruchot, S., Pieri, M. M., Pogge, R., Pollack, J. E., Poppett, C., Prada, F., Prakash, A., Probst, R. G., Rabinowitz, D., Raichoor, A., Ree, C. H., Refregier, A., Regal, X., Reid, B., Reil, K., Rezaie, M., Rockosi, C. M., Roe, N., Ronayette, S., Roodman, A., Ross, A. J., Ross, N. P., Rossi, G., Rozo, E., Ruhlmann-Kleider, V., Rykoff, E. S., Sabiu, C., Samushia, L., Sanchez, E., Sanchez, J., Schlegel, D. J., Schneider, M., Schubnell, M., Secroun, A., Seljak, U., Seo, H.-J., Serrano, S., Shafieloo, A., Shan, H., Sharples, R., Sholl, M. J., Shourt, W. V., Silber, J. H., Silva, D. R., Sirk, M. M., Slosar, A., Smith, A., Smoot, G. F., Som, D., Song, Y.-S., Sprayberry, D., Staten, R., Stefanik, A., Tarle, G., Sien Tie, S., Tinker, J. L., Tojeiro, R., Valdes, F., Valenzuela, O., Valluri, M., Vargas-Magana, M., Verde, L., Walker, A. R., Wang, J., Wang, Y., Weaver, B. A., Weaverdyck, C., Wechsler, R. H., Weinberg, D. H., White, M., Yang, Q., Yeche, C., Zhang, T., Zhao, G.-B., Zheng, Y., Zhou, X., Zhou, Z., Zhu, Y., Zou, H., and Zu, Y. (2016). The DESI Experiment Part I: Science, Targeting, and Survey Design. *arXiv e-prints*, page arXiv:1611.00036. (Citações: 31.)

D'Isanto, A. and Polsterer, K. L. (2017). Photometric redshift estimation via deep learning. *arXiv preprint arXiv:1706.02467*. (Citações: 15, 16, 42, 45 e 46.)

D'Abrusco, R., Staiano, A., Longo, G., Brescia, M., Paolillo, M., De Filippis, E., and Tagliaferri, R. (2007). Mining the sdss archive. i. photometric redshifts in the nearby universe. *The Astrophysical Journal*, 663(2):752. (Citações: 42, 45 e 46.)

D'Isanto, A., Cavuoti, S., Gieseke, F., and Polsterer, K. L. (2018). Return of the features-efficient feature selection and interpretation for photometric redshifts. *Astronomy & Astrophysics*, 616:A97. (Citações: 43, 45 e 46.)

D'Isanto, A. and Polsterer, K. L. (2018). Photometric redshift estimation via deep learning-generalized and pre-classification-less, image based, fully probabilistic redshifts. *Astronomy & Astrophysics*, 609:A111. (Citações: 86 e 99.)

Faes, D. M. (2018). Use of python programming language in astronomy and science. *arXiv preprint arXiv:1807.04806*. (Citações: 56.)

Fukugita, M., Shimasaku, K., Ichikawa, T., Gunn, J., et al. (1996). The sloan digital sky survey photometric system. Technical report, SCAN-9601313. (Citações: 57.)

- Gardner, M. W. and Dorling, S. (1998). Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric environment*, 32(14):2627–2636. (Citações: 34.)
- Glorot, X. and Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. In Teh, Y. W. and Titterton, M., editors, *Understanding the difficulty of training deep feedforward neural networks*, volume 9 of *Proceedings of Machine Learning Research*, pages 249–256, Chia Laguna Resort, Sardinia, Italy. JMLR Workshop and Conference Proceedings. (Citações: 85.)
- Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press. <http://www.deeplearningbook.org>. (Citações: 34.)
- Gothard, L. Q. and Rosen, J. (2010). *Encyclopedia of physical science*. Facts on File. (Citações: 28.)
- Gschwend, J., Rosell, A. C., Ogando, R., Neto, A. F., Maia, M., da Costa, L., Lima, M., Pellegrini, P., Campisano, R., Singulani, C., et al. (2017). Des science portal: I-computing photometric redshifts. *arXiv preprint arXiv:1708.05643*. (Citações: 15.)
- Hornik, K. (1991). Approximation capabilities of multilayer feedforward networks. *Neural networks*, 4(2):251–257. (Citações: 34.)
- Hoyle, B. (2016). Measuring photometric redshifts using galaxy images and deep neural networks. *Astronomy and Computing*, 16:34–40. (Citações: 42, 45 e 46.)
- Huang, G.-B. (2003). Learning capability and storage capacity of two-hidden-layer feedforward networks. *IEEE Transactions on Neural Networks*, 14(2):274–281. (Citações: 84.)
- Hubble, E. (1929). A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae. *Proceedings of the National Academy of Science*, 15(3):168–173. (Citações: 14, 26 e 27.)
- Ivezić, Ž., Connolly, A., Vanderplas, J., and Gray, A. (2014). *Statistics, Data Mining and Machine Learning in Astronomy*. Princeton University Press. (Citações: 25.)
- Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*. (Citações: 85.)

- Kremer, J., Stensbo-Smidt, K., Gieseke, F., Pedersen, K. S., and Igel, C. (2017). Big universe, big data: machine learning and image analysis for astronomy. *IEEE Intelligent Systems*, 32(2):16–22. (Citações: 14.)
- Leavitt, H. S. (1908). 1777 variables in the Magellanic Clouds. *Annals of Harvard College Observatory*, 60:87–108.3. (Citações: 26.)
- Lehmann, E. L. and Casella, G. (2006). *Theory of point estimation*. Springer Science & Business Media. (Citações: 104.)
- Mickaelian, A. M. (01 Mar. 2016). Astronomical surveys and big data. *Open Astronomy*, 25(1):75 – 88. (Citações: 30.)
- Reis, R. R., Soares-Santos, M., Annis, J., Dodelson, S., Hao, J., Johnston, D., Kubo, J., Lin, H., Seo, H.-J., and Simet, M. (2012). The sloan digital sky survey co-add: A galaxy photometric redshift catalog. *The Astrophysical Journal*, 747(1):59. (Citações: 15.)
- Rokach, L. and Maimon, O. (2015). *Data Mining With Decision Trees: Theory And Applications (2Nd Edition)*. World Scientific Publishing Company. (Citações: 36.)
- Sánchez, C., Carrasco Kind, M., Lin, H., Miquel, R., Abdalla, F. B., Amara, A., Banerji, M., Bonnett, C., Brunner, R., Capozzi, D., et al. (2014). Photometric redshift analysis in the dark energy survey science verification data. *Monthly Notices of the Royal Astronomical Society*, 445(2):1482–1506. (Citações: 16, 42, 45, 46 e 104.)
- Sautermeister, B. (2016). Deep Learning Approaches to Predict Future Frames in Videos. Master's thesis, Technische Universitat Munchen, Germany. (Citações: 33.)
- Singh, S. and Dayan, P. (1998). Analytical mean squared error curves for temporal difference learning. *Machine Learning*, 32(1):5–40. (Citações: 104.)
- Soumagnac, M. et al. (2015). Star/Galaxy Separation at Faint Magnitudes: Application to a Simulated Dark Energy Survey. *Mon. Not. Roy. Astron. Soc.*, 450(1):666–680. (Citações: 102.)
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958. (Citações: 16, 49 e 100.)

- Stevens, E. and Antiga, L. (2019). *Deep Learning with PyTorch: Essential Excerpts*. Manning Publications. <https://pytorch.org/assets/deep-learning/Deep-Learning-with-PyTorch.pdf>. (Citações: 33 e 35.)
- Sullivan, W. (2017). *Machine Learning For Beginners: Algorithms, Decision Tree & Random Forest Introduction*. Healthy Pragmatic Solutions Inc. (Citações: 36.)
- Syarifudin, M., Hakim, M., and Arifyanto, M. (2019). Applying deep neural networks (dnn) for measuring photometric redshifts from galaxy images: Preliminary study. In *Journal of Physics: Conference Series*, volume 1231, page 012013. IOP Publishing. (Citações: 43, 45 e 46.)
- Thompson, D. R., Burke-Spolaor, S., Deller, A. T., Majid, W. A., Palaniswamy, D., Tingay, S. J., Wagstaff, K. L., and Wayth, R. B. (2014). Real-time adaptive event detection in astronomical data streams. *IEEE Intelligent Systems*, 29(1):48–55. (Citações: 14.)
- Way, M. J. and Klose, C. (2012). Can self-organizing maps accurately predict photometric redshifts? *Publications of the Astronomical Society of the Pacific*, 124(913):274. (Citações: 104.)
- Willmott, C. J. and Matsuura, K. (2005). Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance. *Climate research*, 30(1):79–82. (Citações: 105.)