

SHAPE TIME SERIES DAY



CEFET/RJ

CONCEPT DRIFT

Eduardo Ogasawara

eogasawara@ieee.org

<https://eic.cefet-rj.br/~eogasawara>

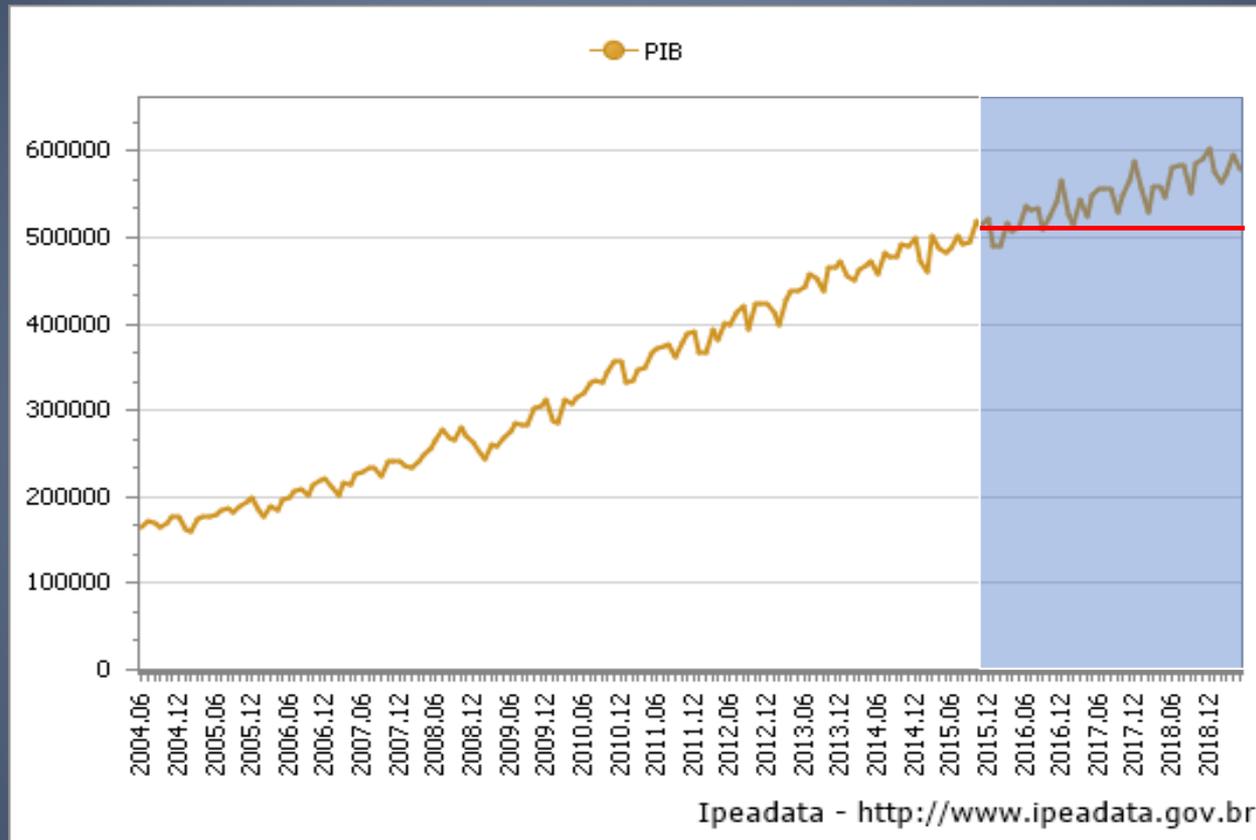
Leonardo Carvalho

lm.carvalho@petrobras.com.br

Não-estacionariedade

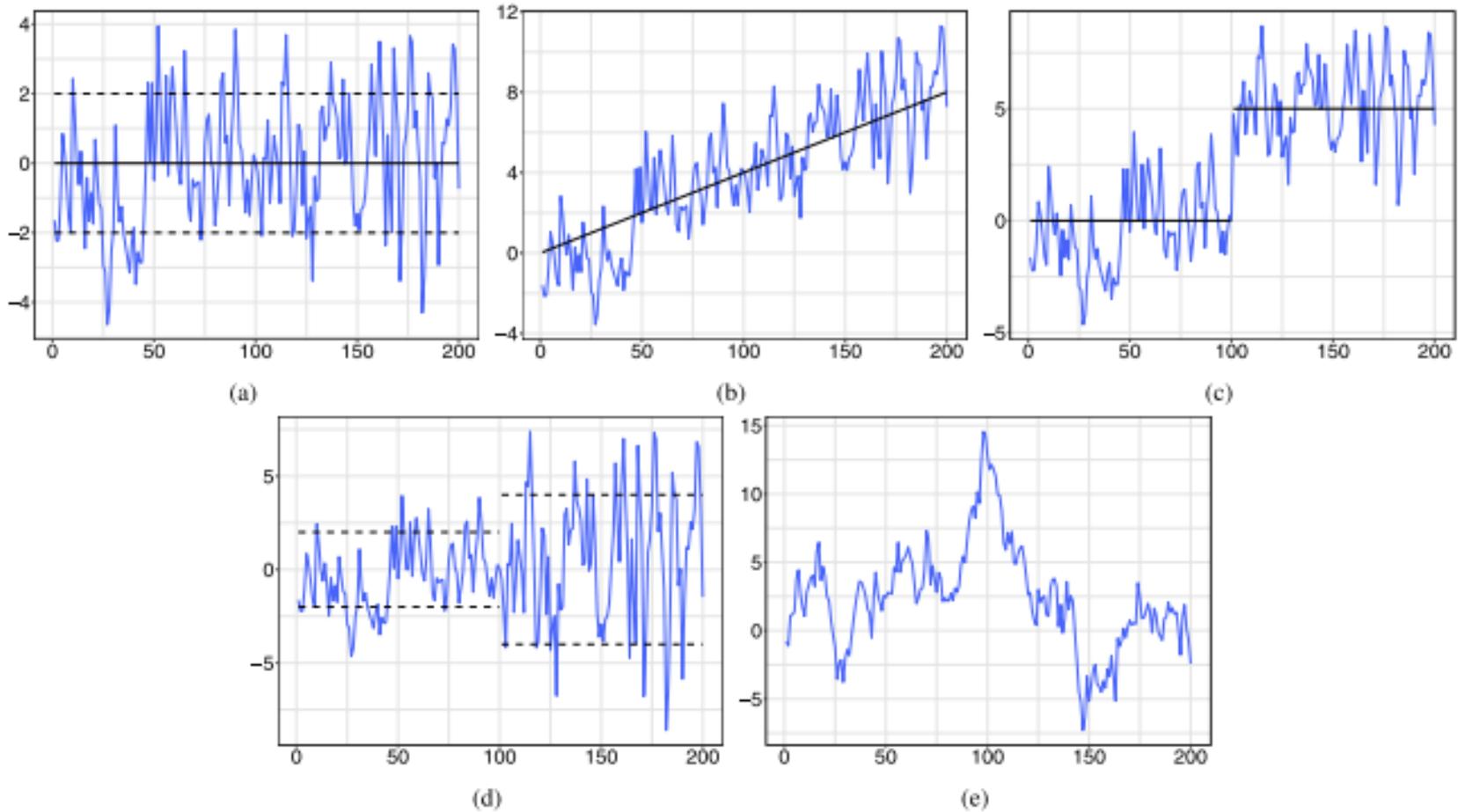
- Estacionariedade
 - Dataset D
 - Amostras D_s
 - Propriedades estatísticas em D_s não variam com o tempo
 - Séries temporais: média, variância e covariância
- Não-estacionariedade
- Métodos de data analytics
 - A grande maioria dos métodos assumem “implicitamente” estacionariedade

Predição de séries temporais



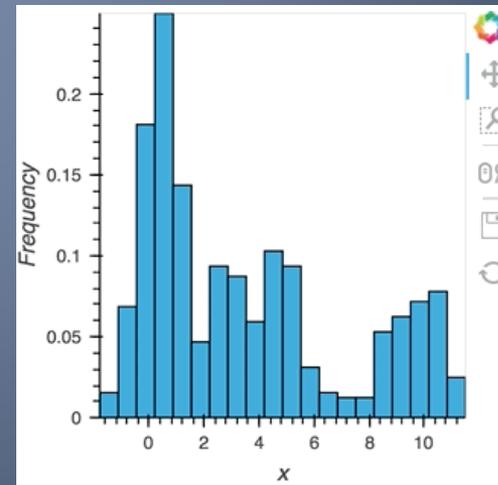
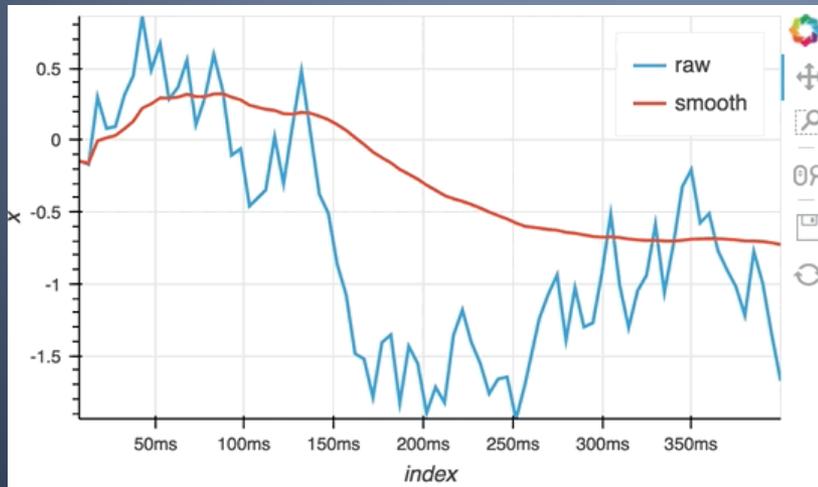
Normalização (min/max)

Tipos de não-estacionariedade



As múltiplas faces da não-estacionariedade

- Crítica de Lucas
- Pseudo-estacionariedade
- Dilema da Plasticidade e Estabilidade
- Padrões emergentes (*Emerging patterns*)
- Mudança de conceito (*Concept Drift*)



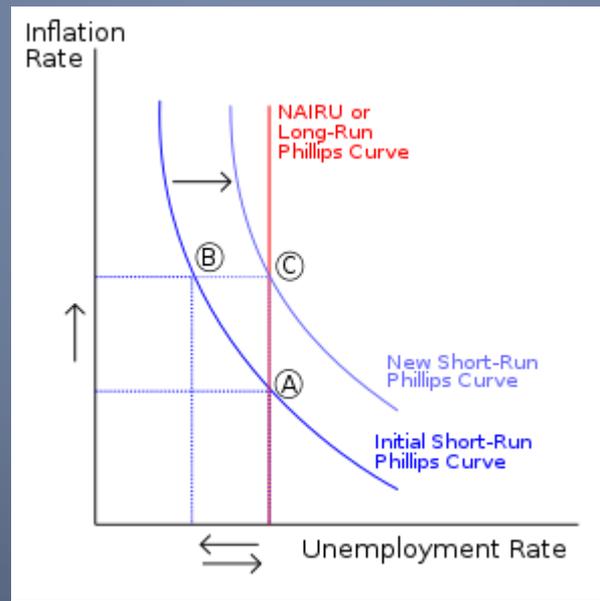
[1] A.S. Iwashita and J.P. Papa, 2019, An Overview on Concept Drift Learning, *IEEE Access*, v. 7, p. 1532–1547.

[2] G. Ditzler, M. Roveri, C. Alippi, and R. Polikar, 2015, Learning in Nonstationary Environments: A Survey, *IEEE Computational Intelligence Magazine*, v. 10, n. 4, p. 12–25.

[3] D. Gujarati, 2002, *Basic Econometrics*. 4 ed. Boston; Montreal, McGraw-Hill/Irwin. / [4] S.O. Haykin, 2008, *Neural Networks and Learning Machines*. 3 ed. New York, Prentice Hall.

Crítica de Lucas

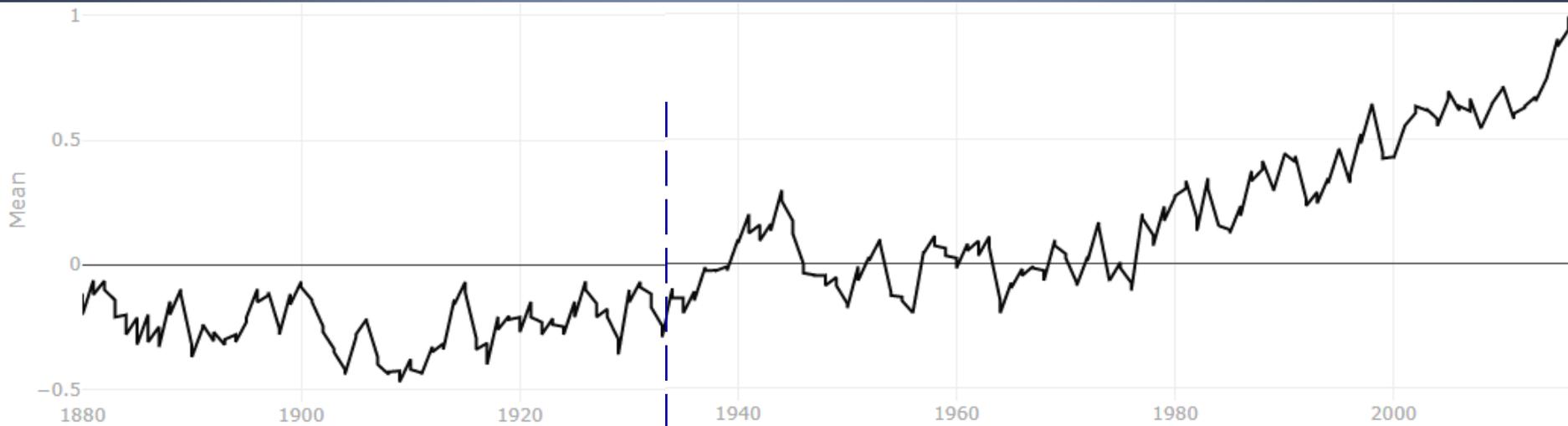
- “Dado que a estrutura de um modelo econométrico consiste em regras de decisão ótimas dos agentes econômicos, e que as regras de decisão ótimas variam sistematicamente com as mudanças na estrutura das séries relevantes para o decisor, conclui-se que qualquer mudança na política sistematicamente irá alterar a estrutura dos modelos econométricos”



Pseudo-estacionariedade

Monthly mean temperature anomalies in degrees Celsius relative to a base period

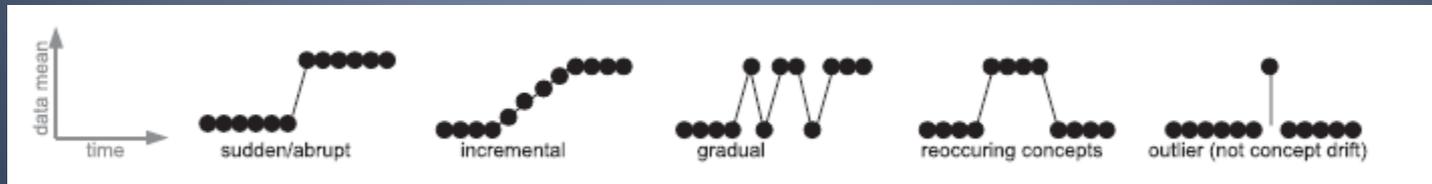
Período pseudo-estacionário



Ponto de mudança

Dilema da Plasticidade e Estabilidade

- Redes neurais são conhecidas pela adaptabilidade
 - Capacidade de atualizar os pesos em função de alterações no ambiente
 - Treinamento incremental
 - Alteração dos pesos sinápticos
- Sistemas adaptativos visam abordar não-estacionariedade
 - Buscando-se robustez, adota-se adaptabilidade
 - Maior adaptabilidade, mais suscetível a situações espúrias, menor robustez
 - Dilema: encontrar o tempo certo para se adaptar



[1] S.O. Haykin, 2008, *Neural Networks and Learning Machines*. 3 ed. New York, Prentice Hall.

[2] Grossberg, S., 1988. *Neural Networks and Natural Intelligence*, Cambridge, MA: MIT Press.

[3] G. Ditzler, M. Roveri, C. Alippi, e R. Polikar, 2015, Learning in Nonstationary Environments: A Survey, *IEEE Computational Intelligence Magazine*, v. 10, n. 4, p. 12–25

Emerging patterns

- Padrões emergentes são coleções de itens cuja frequência muda de um dataset (batch) para outro
- Datasets D_t (anterior) e D_u (próximo)
 - Crescimento para itens χ : $\rho(\chi)$

$$\rho(\chi) = \begin{cases} \infty, & \text{support}_t(i) = 0 \\ 0, & \text{support}_t(\chi) = \text{support}_u(\chi) = 0 \\ \frac{\text{support}_u(\chi)}{\text{support}_t(\chi)}, & \text{otherwise} \end{cases}$$

- Dado um limite σ , um padrão χ é emergente se $\rho(\chi) \geq \sigma$

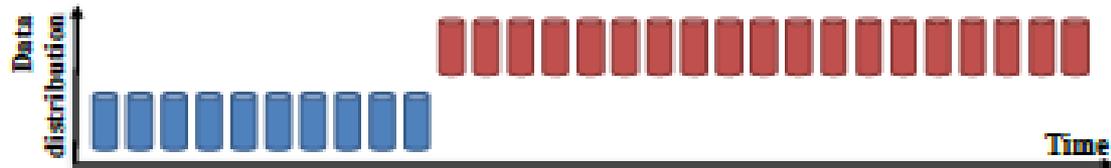
Concept Drift

- Aprendizado no contexto de distribuições não-estacionárias
 - Aprendizado é feito em lotes (batches)
 - Data streams (objetos com timestamps)
 - Definições
 - $P(Y)$ probabilidade da variável dependente (rótulo)
 - $P(X)$ probabilidade das variáveis independentes (objetos)
 - $P(X, Y)$ probabilidade conjunta dos objetos e rótulo
 - $P(Y|X)$ distribuição provável do rótulo para objeto
 - Concept = $P(X, Y) = P(\chi)$
 - Drift = $P_t(\chi) \neq P_u(\chi)$

Tipos de Concept Drift

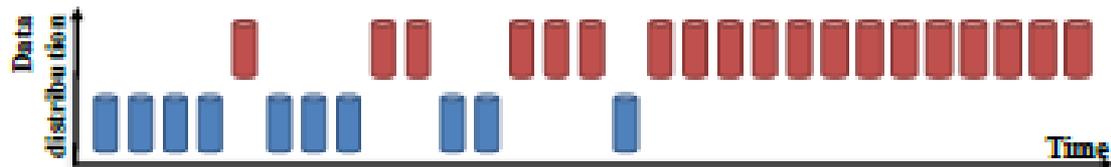
Sudden Drift:

A new concept occurs within a short time.



Gradual Drift:

A new concept gradually replaces an old one over a period of time.



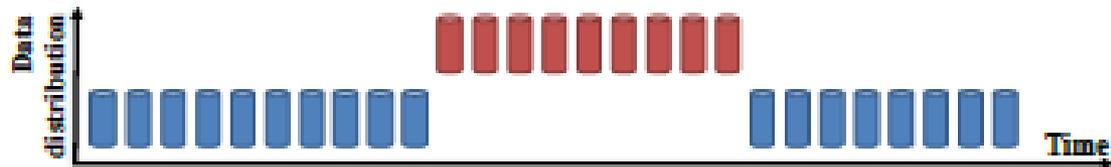
Incremental Drift:

An old concept incrementally changes to a new concept over a period of time.



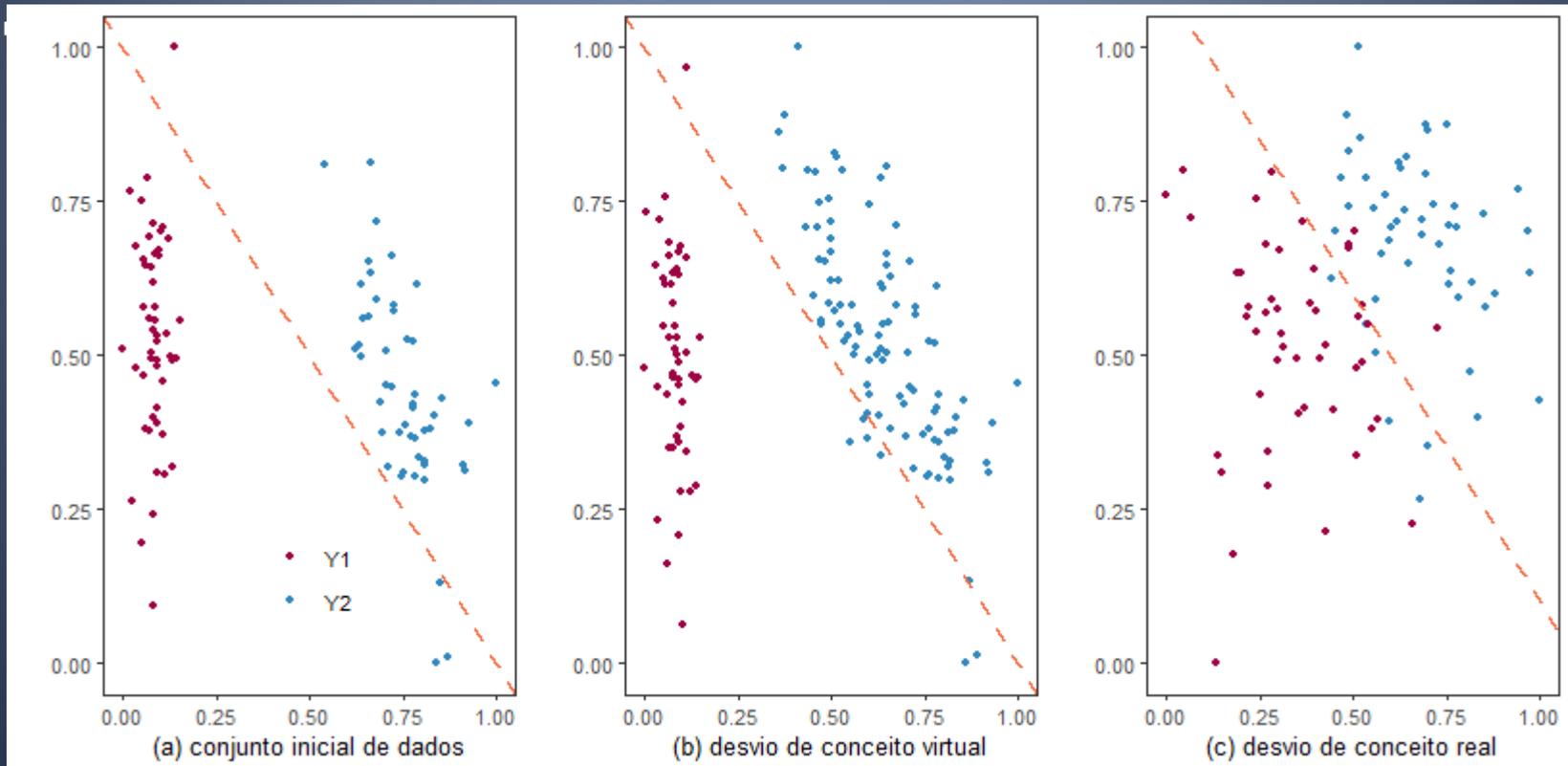
Reoccurring Concepts:

An old concept may reoccur after some time.



Magnitude e Real/Virtual Concept Drift

- Magnitude do Concept Drift: $D(t, u)$
- Real Concept Drift
 - $P_t(Y|X) \neq P_u(Y|X)$ e $P_t(X) = P_u(X)$
- Virtual Concept Drift



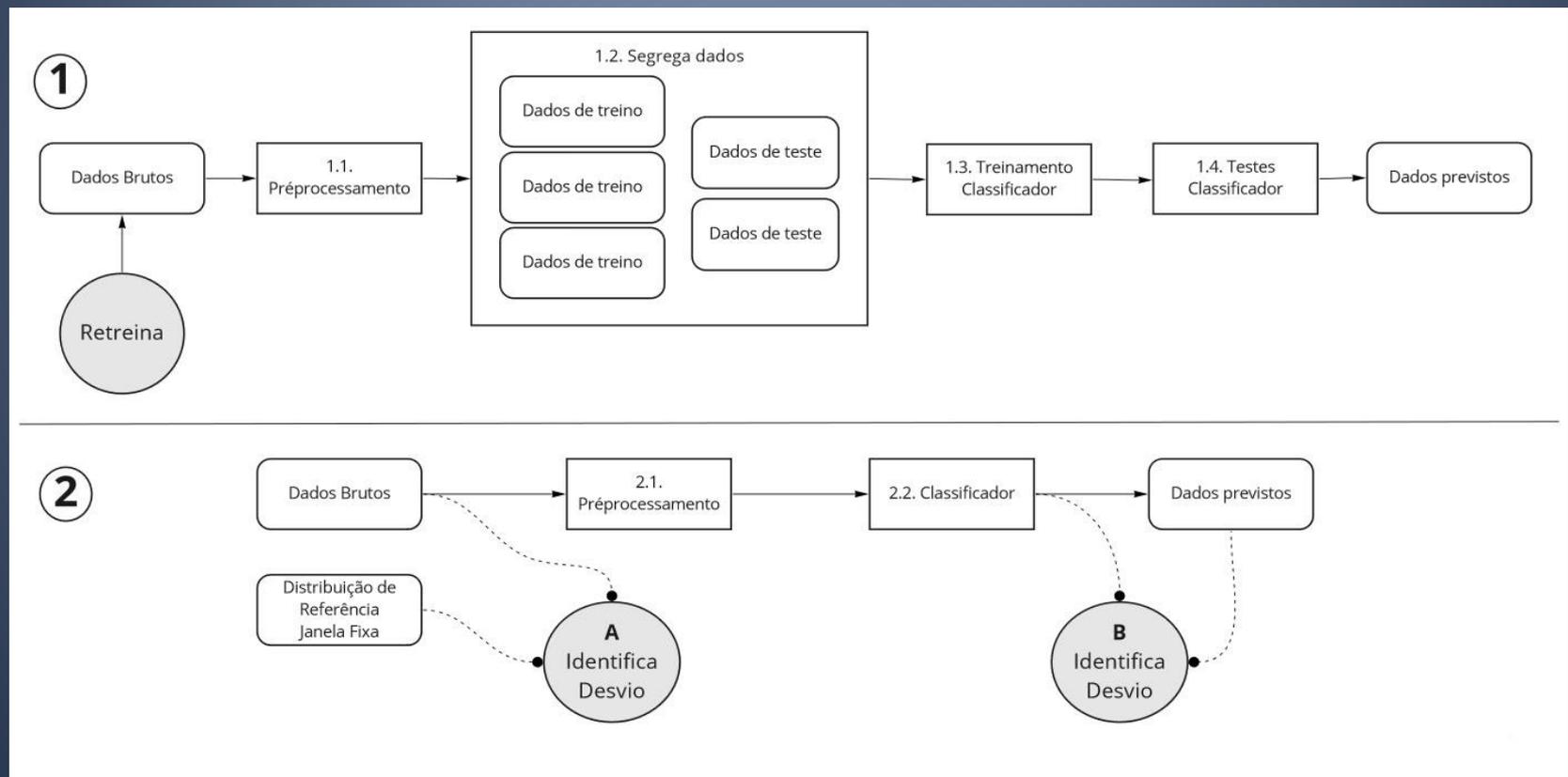
Taxonomia de não-estacionariedade



Detecção e tratamento de desvio de conceito

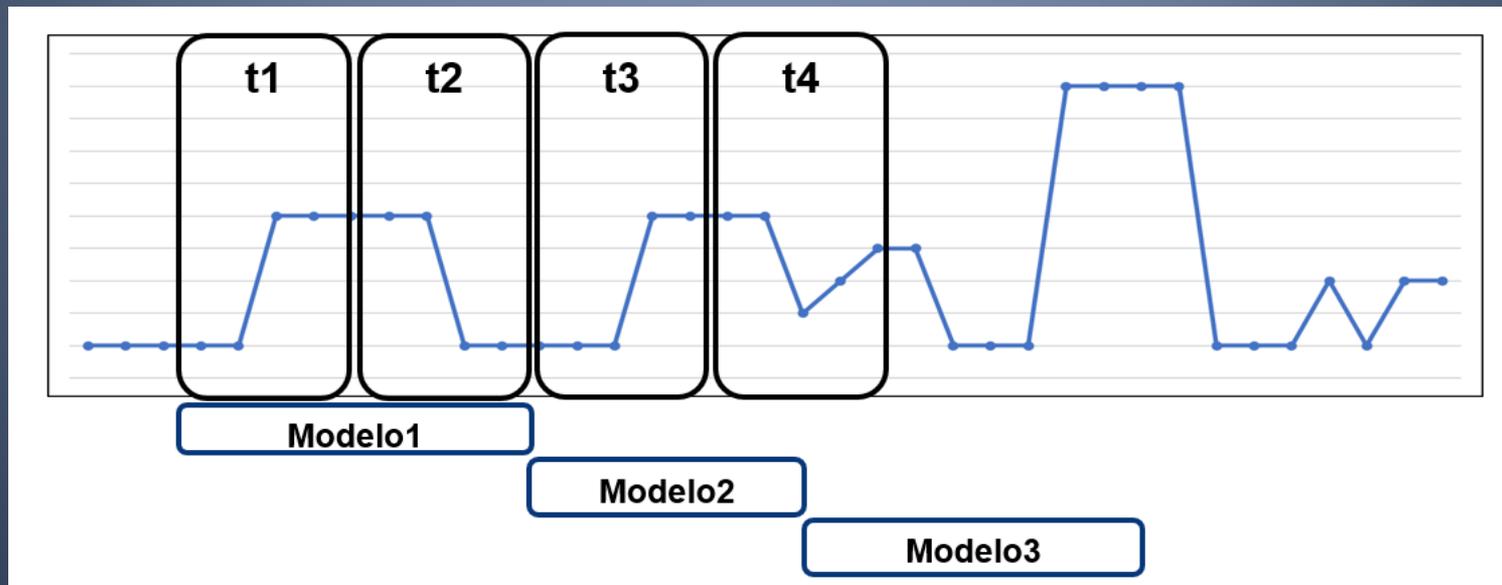
Desvio de Conceito – Detecção

- Ponto A – identificação do desvio diante da comparação das distribuições
- Ponto B – identificação do desvio a partir da evolução de algum indicador de desempenho

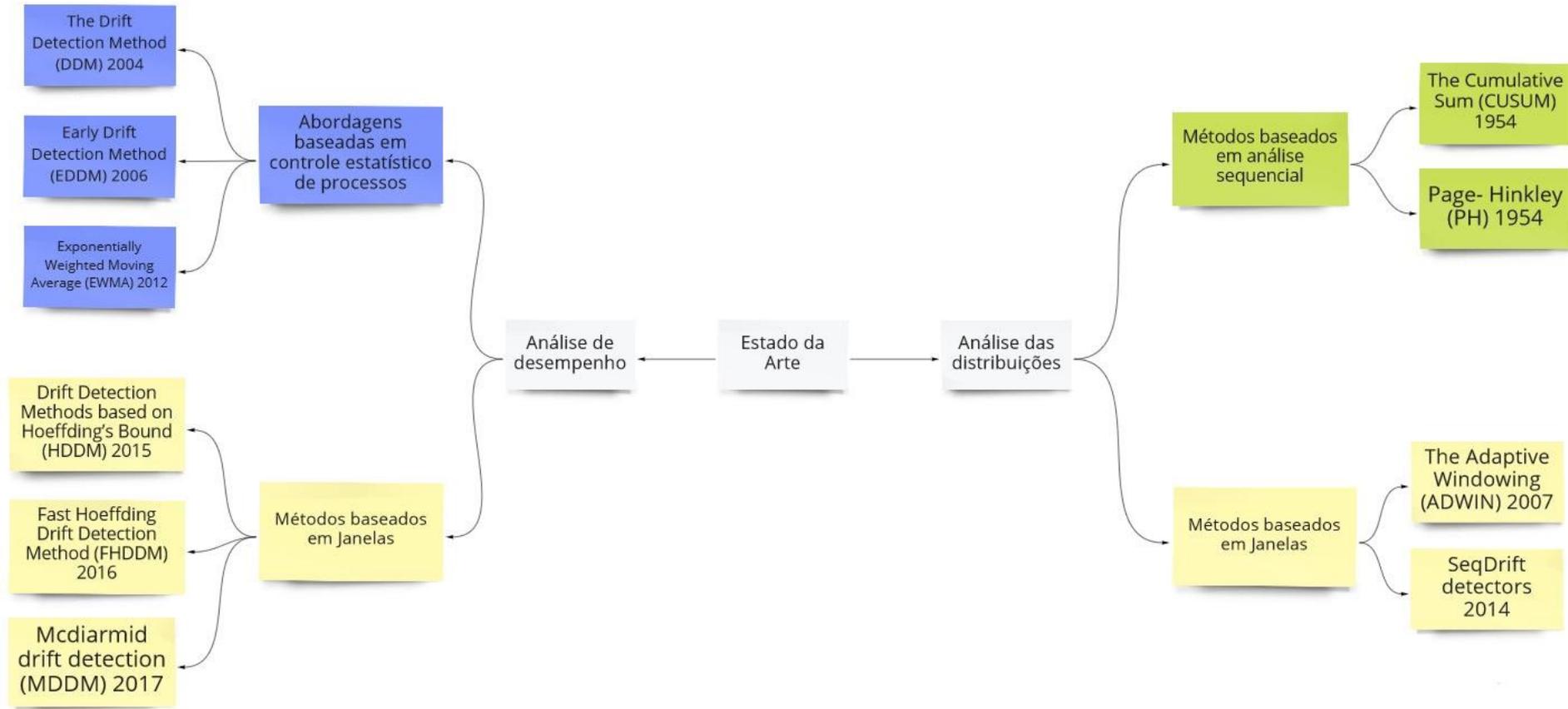


Desvio de Conceito – Tratamento

- Passivo – Métodos que adaptam o método em intervalos regulares, sem considerar se as mudanças realmente ocorreram.
- Ativo – Métodos que alteram apenas o modelo de classificação após a detecção de uma alteração. Eles são usados em conjunto com um modelo de detecção.



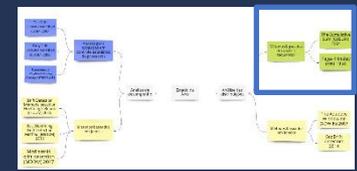
Detecção de Desvio de Conceito – Classificação e Estado da Arte



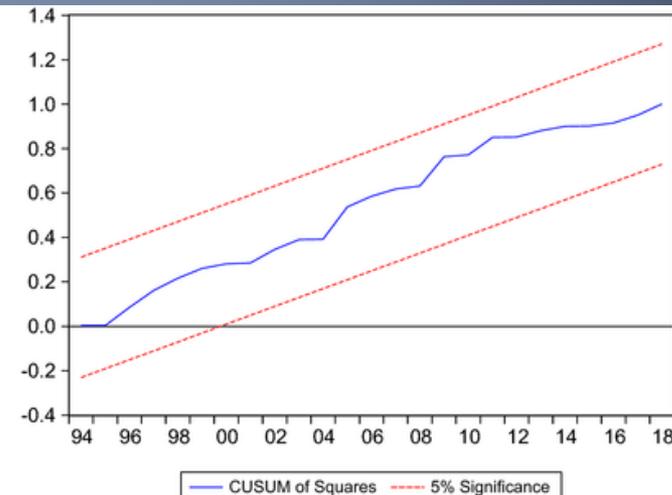
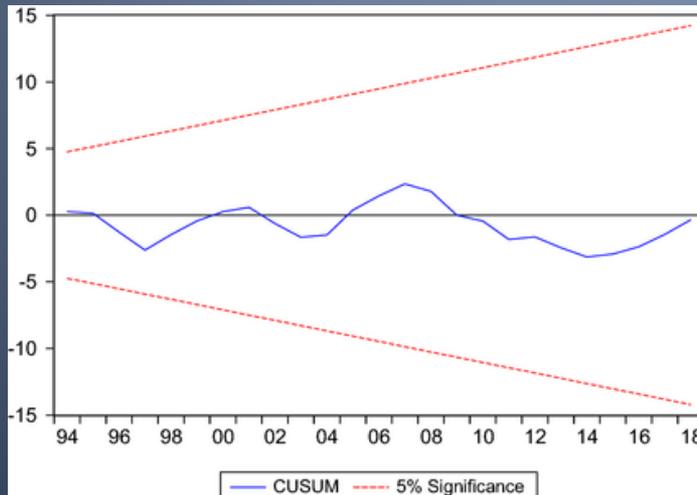
Taxonomia adaptada de Gama et al. [2014] e Pesaranghader et al. [2016].

Análise das Distribuições

Métodos Baseados em Análise Sequencial

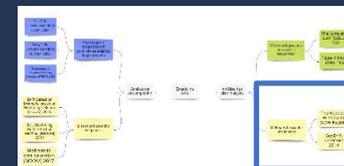


- Métodos baseados em análise sequencial avaliam sequencialmente os resultados de previsão à medida que se tornam disponíveis
- Quando a diferença da predição e o valor é maior que o limite definido, sinaliza-se um desvio. O limite depende da taxa admissível de falsos positivos
- Neste grupo estão os métodos Cumulative Sum (CUSUM) e sua variante Page-Hinkley (PH), que trata médias zeradas [PAGE, 1954]



Análise das Distribuições

Métodos Baseados em Janelas



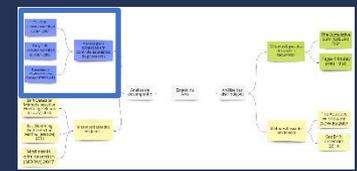
- Utiliza-se uma janela de referência fixa para resumir as informações anteriores e uma janela deslizante para resumir as informações mais recentes
 - Uma diferença significativa entre as distribuições dessas duas janelas implica na ocorrência de um desvio
- **ADWIN** (Adaptive Sliding Window)- analisa a média entre as duas janelas
- **SeqDrift** – Analisa a soma entre as duas janelas

ADWIN: ADAPTIVE WINDOWING ALGORITHM

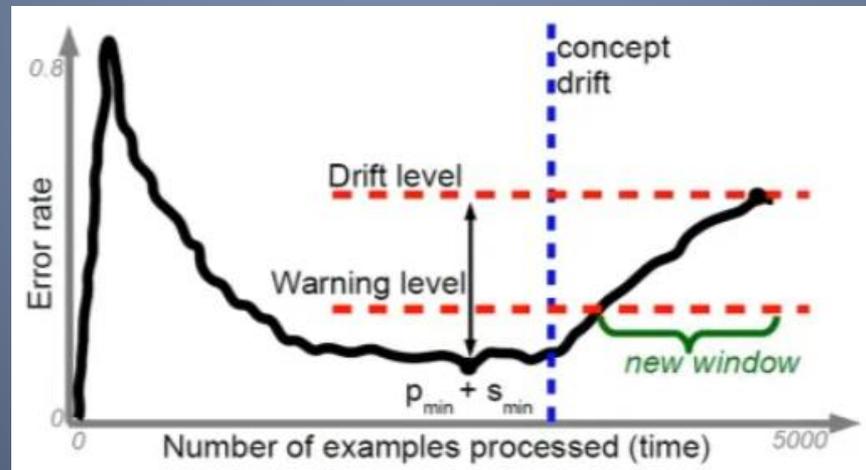
```
1 Initialize Window  $W$ 
2 for each  $t > 0$ 
3   do  $W \leftarrow W \cup \{x_t\}$  (i.e., add  $x_t$  to the head of  $W$ )
4   repeat Drop elements from the tail of  $W$ 
5     until  $|\hat{\mu}_{W_0} - \hat{\mu}_{W_1}| \geq \epsilon_{cut}$  holds
6     for every split of  $W$  into  $W = W_0 \cdot W_1$ 
7   output  $\hat{\mu}_W$ 
```

Análise de Desempenho

Controle Estatístico de Processos

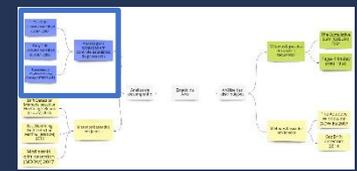


- Analisa parâmetros estatísticos, como a média e o desvio padrão associados a taxa de erros, a fim de detectar o desvio
- **DDM** (Drift Detection Method) - Calcula a probabilidade de classificação incorreta (p_i) e seu desvio padrão (σ_i) para cada instância. Analisa se a taxa de erro aumenta

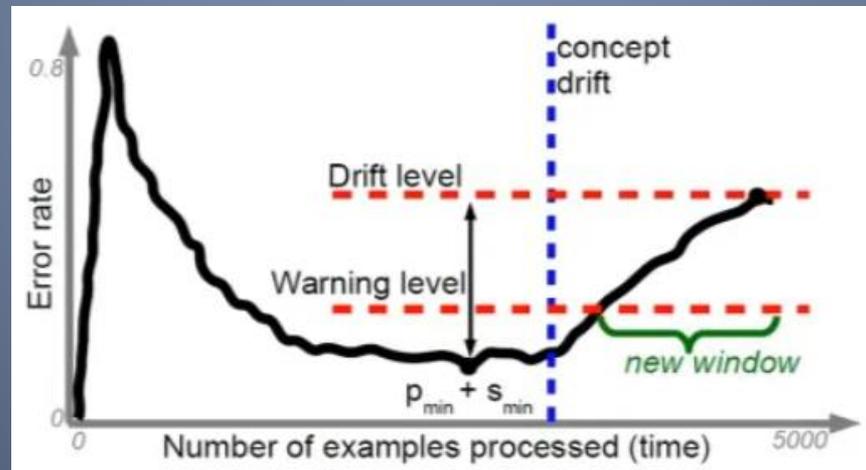


Análise de Desempenho

Controle Estatístico de Processos

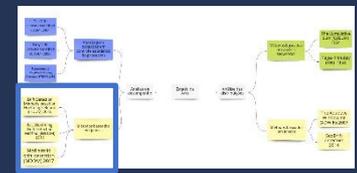


- **EDDM** (Early Drift Detection Method) - O mesmo mecanismo de alerta-alarme do DDM, mas em vez de usar a taxa de erro do classificador, foi proposta a distância entre erros
 - Limite do que seja considerado desvio considera proporção entre média e desvio padrão correntes e a média máxima e duas vezes o desvio padrão máximo

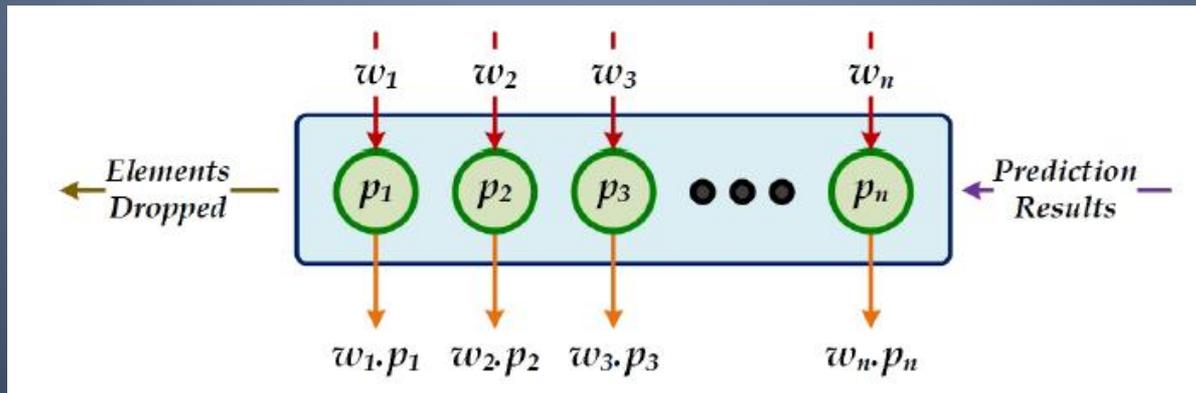


Análise de Desempenho

Métodos Baseados em Janelas

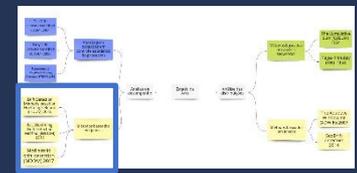


- **HDDM** (Drift Detection Method based on the Hoeffding's inequality) – Desvio de conceito baseados em limites de Hoeffding. Sua mecânica funciona através da monitoração de uma média estimada de desempenho, a partir de erros e acertos de um classificador, e então detectar mudanças relevantes. Utiliza desigualdade de Hoeffding delimitar o que seria o desvio. [FRÍAS-BLANCO et al., 2015]
- **MDDM** (McDiarmid Drift Detection Method) - mantém conceito muito parecido com o HDDM, porém, utiliza a desigualdade de McDiarmid para limite de detecção de desvio. Tipos MDDM-A (peso usando esquema aritmético), MDDM-G (peso usando esquema geométrico) e MDDM-E (peso usando esquema de Euler) [PESARANGHADER et al., 2017].

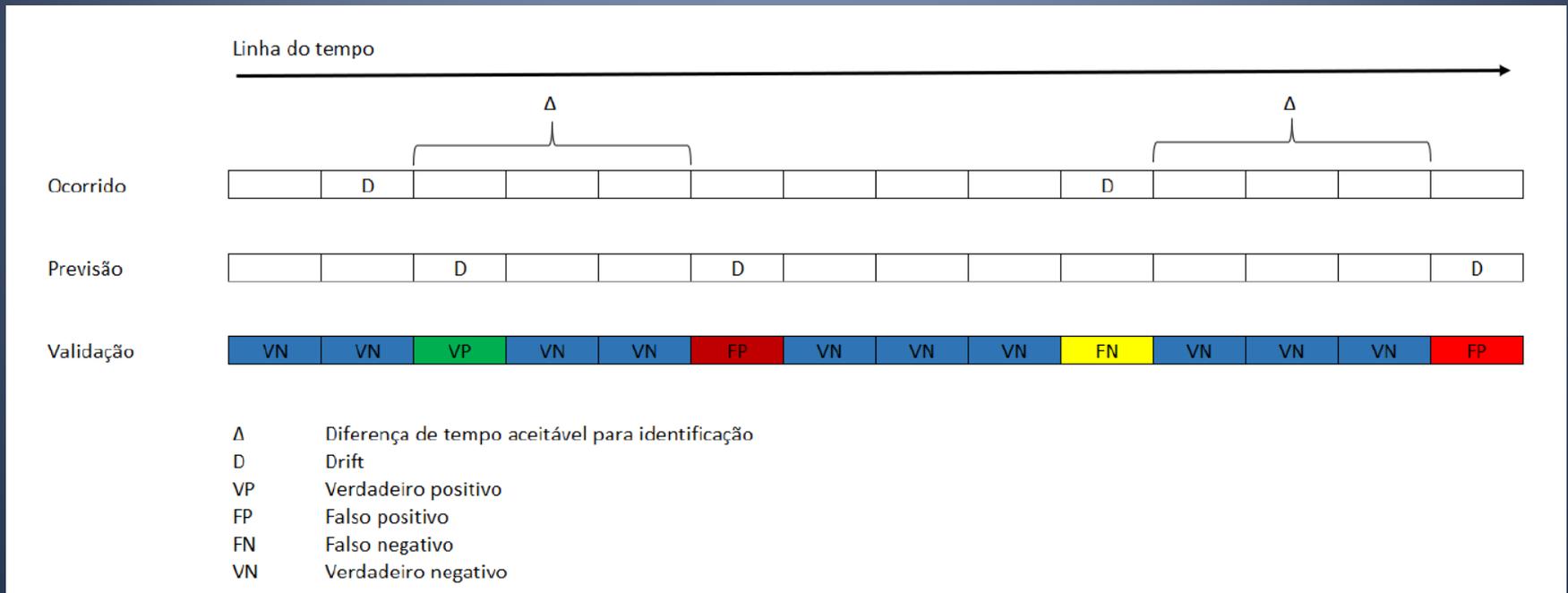


Análise de Desempenho

Métodos Baseados em Janelas

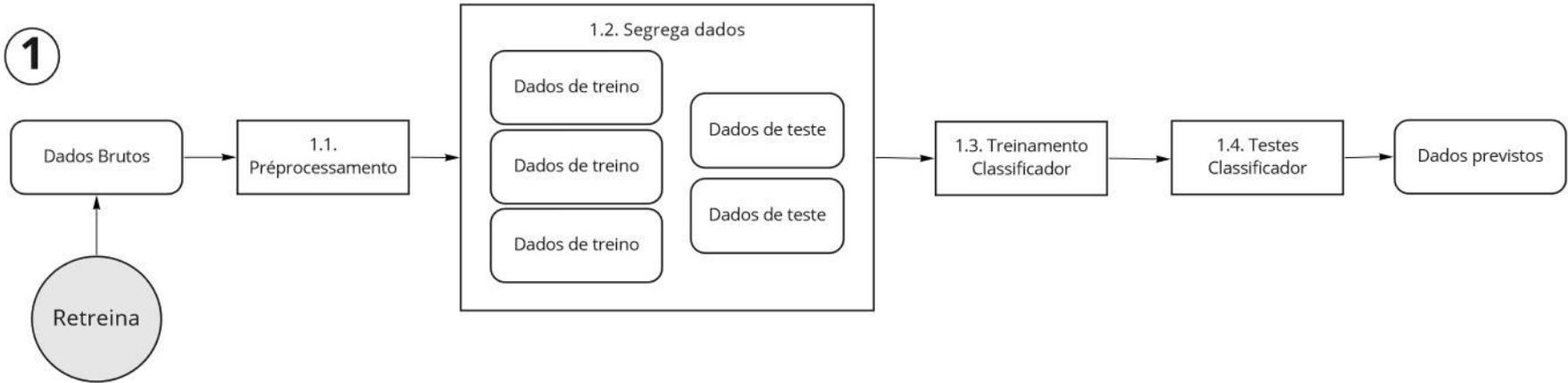


- FHDDM** (Fast Hoeffding Drift Detection Method) – Detecta desvio quando uma diferença significativa entre a probabilidade máxima de predições corretas e a probabilidade mais recente de predições corretas é observada

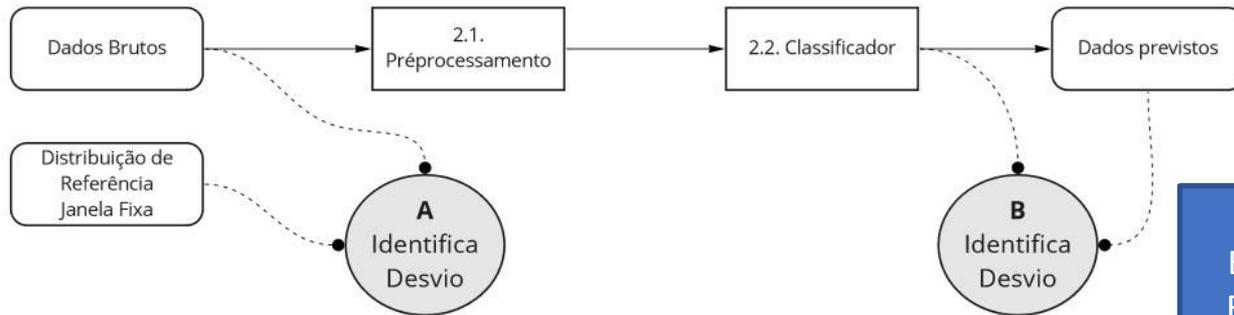


Desvio de Conceito – Classificação dos Métodos

1



2



Page-Hinkley (PH)
CUSUM
ADWIN
SEQDRIFT

DDM
EDDM
RDDM
EWMA
HDDM
MDDM
FHDDM

Modelos em cenário de concept-drift

Abordagens

- Gerência da memória
- Adaptabilidade
- Transformações
- Seleção de modelos

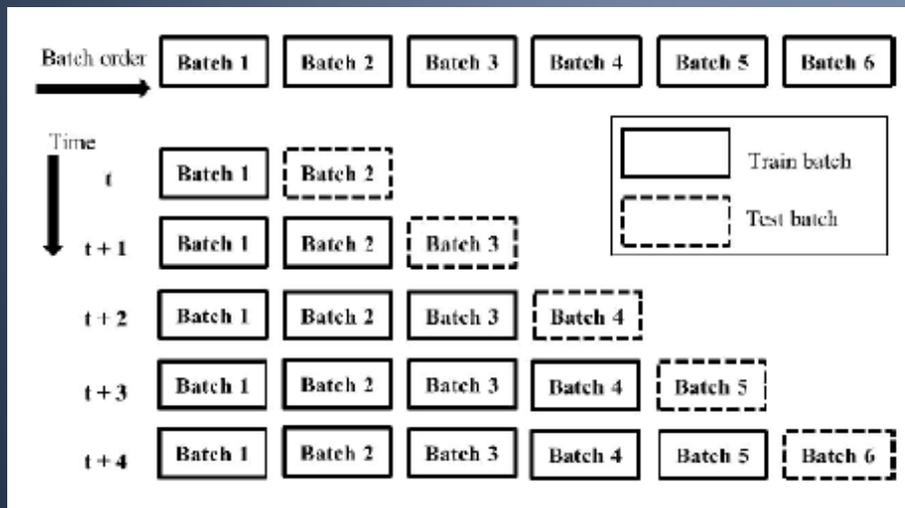
[1] J. Gama, I. Zliobaite, A. Bifet, M. Pechenizkiy, e A. Bouchachia, 2014, A survey on concept drift adaptation, *ACM Computing Surveys*, v. 46, n. 4

[2] A.M. García-Vico, C.J. Carmona, D. Martín, M. García-Borroto, e M.J. del Jesus, 2018, An overview of emerging pattern mining in supervised descriptive rule discovery: taxonomy, empirical study, trends, and prospects, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, v. 8, n. 1

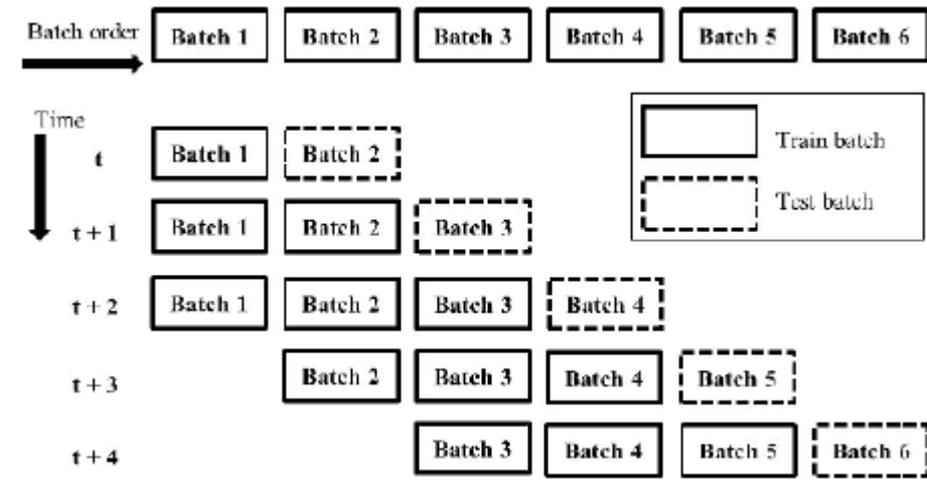
[3] R. Salles, K. Belloze, F. Porto, P. H. Gonzalez, e E. Ogasawara, "Nonstationary time series transformation methods: An experimental review", *Knowledge-Based Systems*, nov. 2018.

Memória

- Processo
 - Testa-se no último batch (previsão)
 - Incorpora-se último batch no treino
- Memória
 - Completa
 - Sem memória
 - Janelas deslizantes



Completa



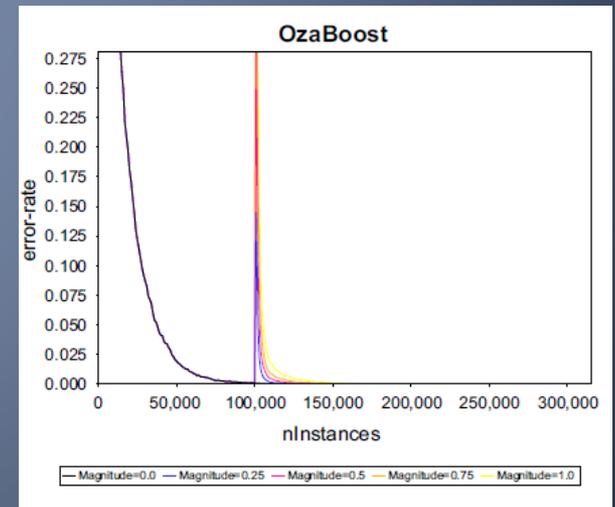
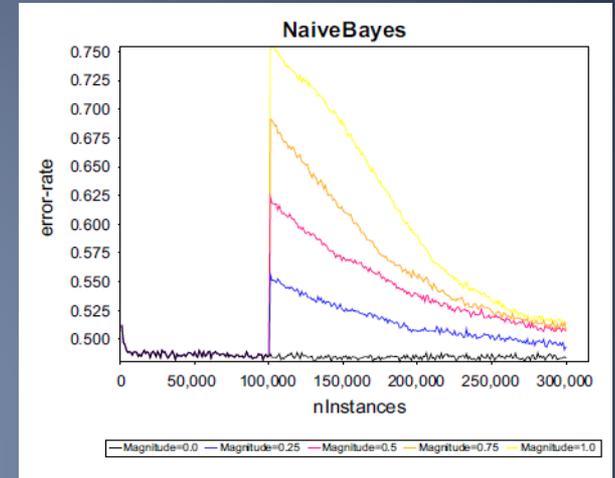
Janela de tamanho 3

[1] J. Gama, I. Zliobaite, A. Bifet, M. Pechenizkiy, e A. Bouchachia, 2014, A survey on concept drift adaptation, *ACM Computing Surveys*, v. 46, n. 4

[2] A.M. García-Vico, C.J. Carmona, D. Martín, M. García-Borroto, e M.J. del Jesus, 2018, An overview of emerging pattern mining in supervised descriptive rule discovery: taxonomy, empirical study, trends, and prospects, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, v. 8, n. 1

Adaptabilidade

- Detecção de *drift*
 - Ativa
 - Passiva
- Aprendizado
 - Incremental
 - Não-incremental
- Modelos
 - Único
 - Ensemble

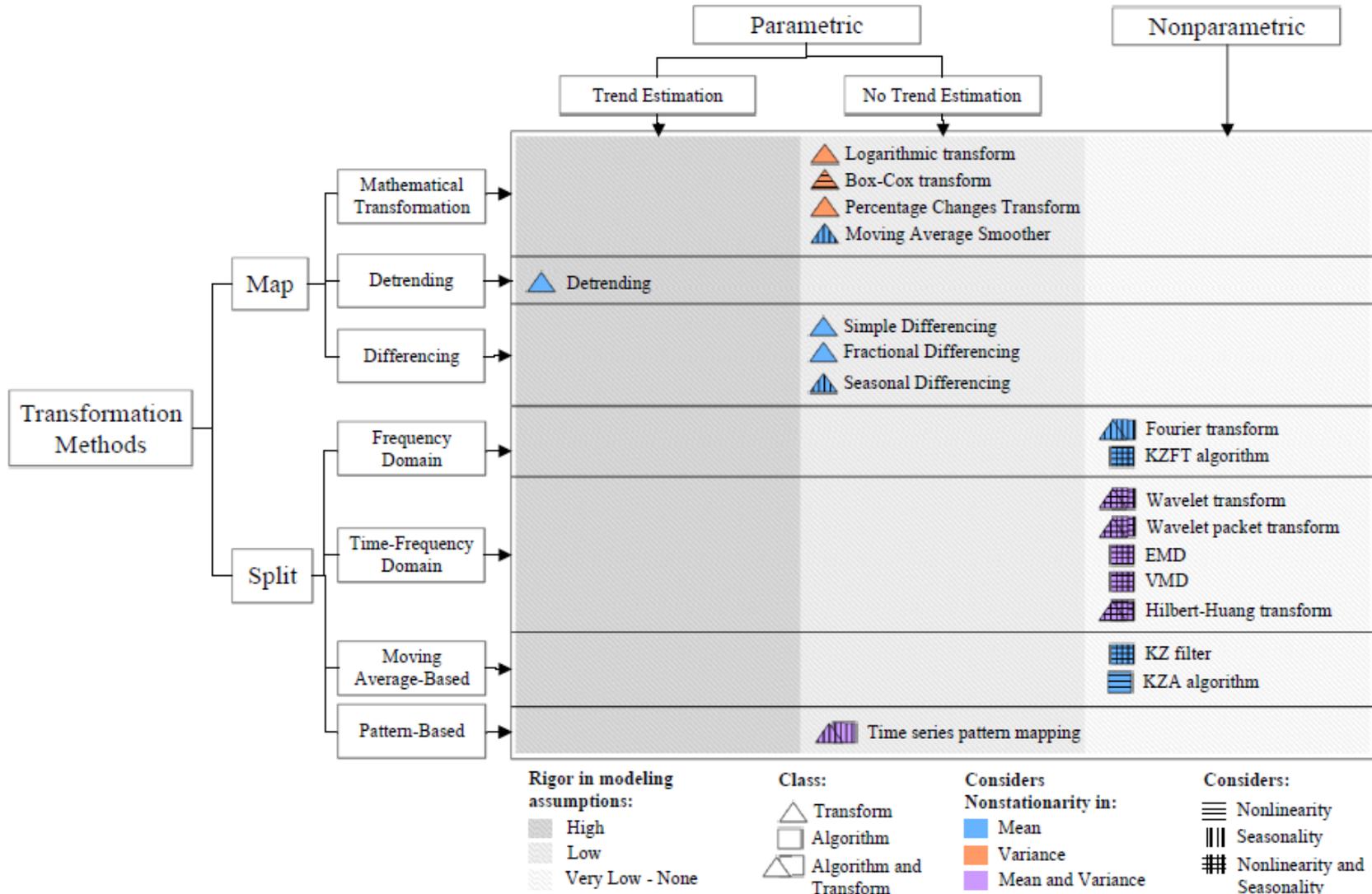


[1] J. Gama, I. Zliobaite, A. Bifet, M. Pechenizkiy, e A. Bouchachia, 2014, A survey on concept drift adaptation, *ACM Computing Surveys*, v. 46, n. 4

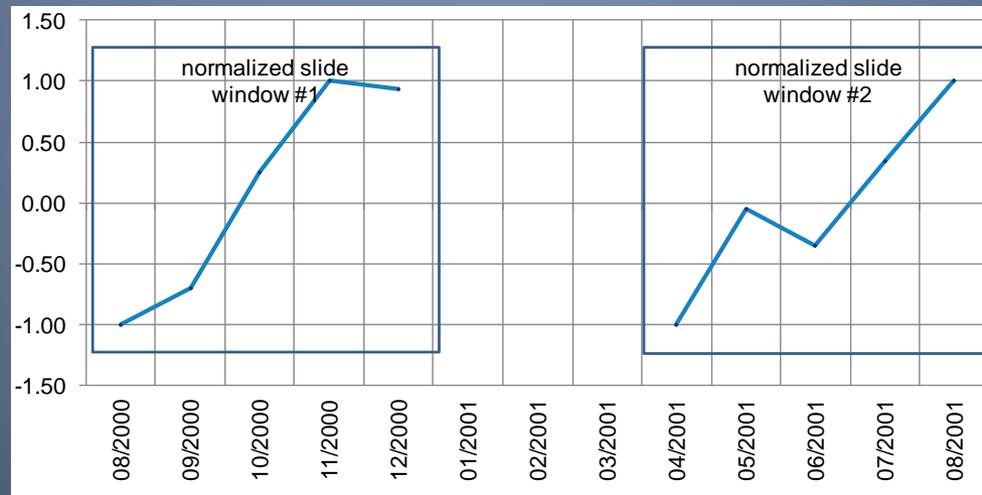
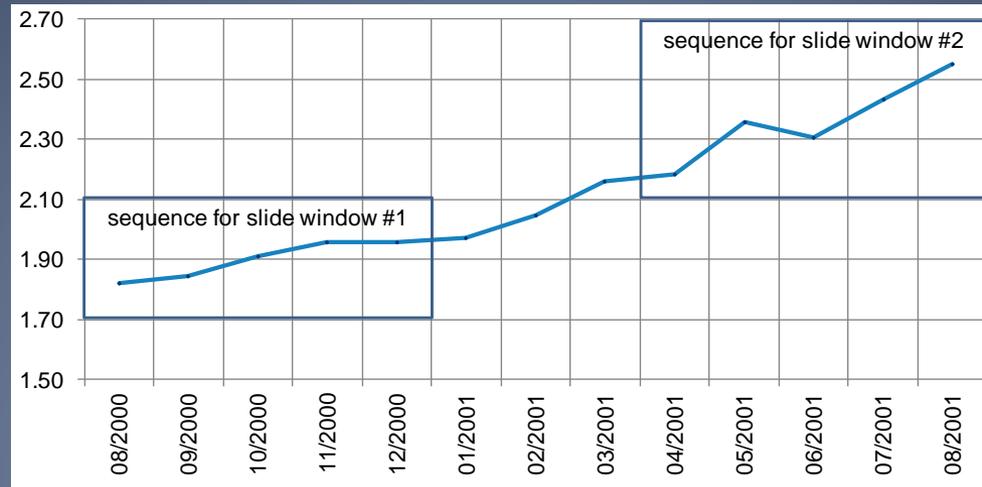
[2] A.M. García-Vico, C.J. Carmona, D. Martín, M. García-Borroto, e M.J. del Jesus, 2018, An overview of emerging pattern mining in supervised descriptive rule discovery: taxonomy, empirical study, trends, and prospects, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, v. 8, n. 1

[3] G.I. Webb, R. Hyde, H. Cao, H.L. Nguyen, e F. Petitjean, 2016, Characterizing concept drift, *Data Mining and Knowledge Discovery*, v. 30, n. 4, p. 964–994.

Transformações



Problemas de normalização usando janelas deslizantes



Monthly average exchange rate of U.S. Dollar to Brazilian Real normalized by sliding window technique from aug/2000 to dec/2000 and from apr/2001 to aug/2001

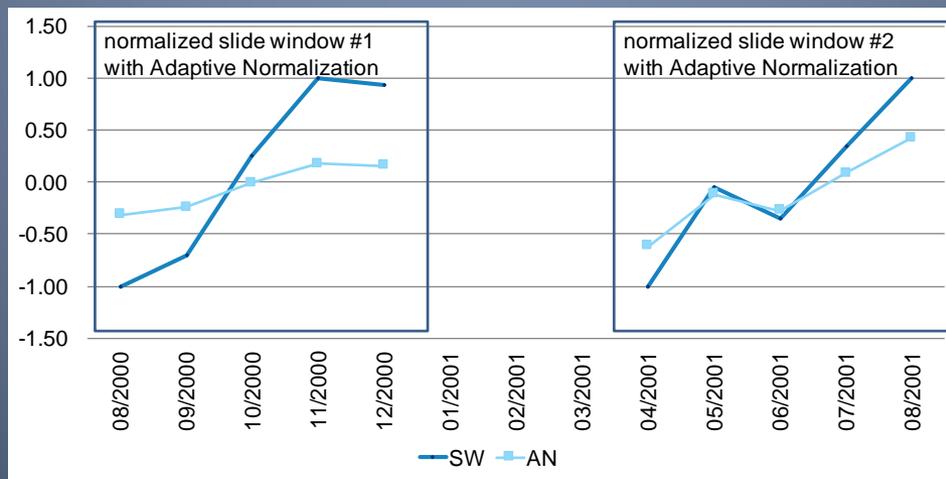
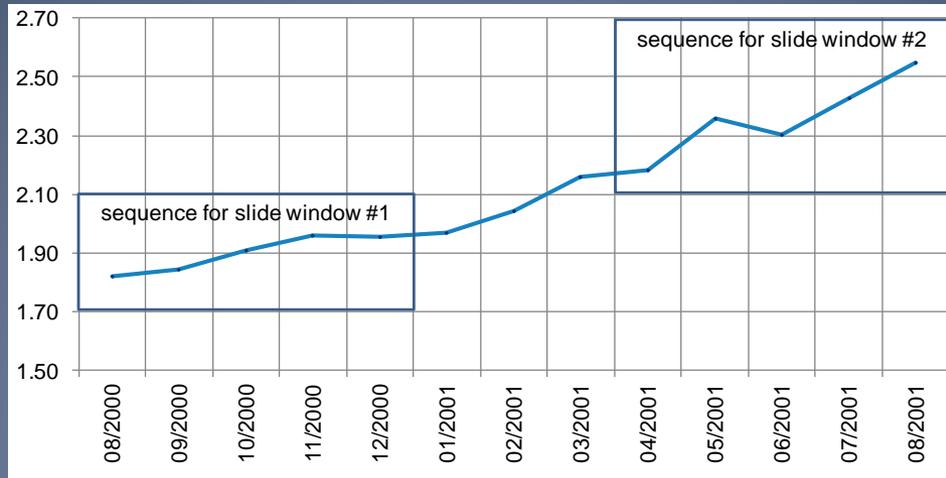
[1] E. Ogasawara, L. Murta, G. Zimbrão, and M. Mattoso, 2009, Neural networks cartridges for data mining on time series, In: *Proceedings of the International Joint Conference on Neural Networks*, p. 2302–2309

Funções inerciais

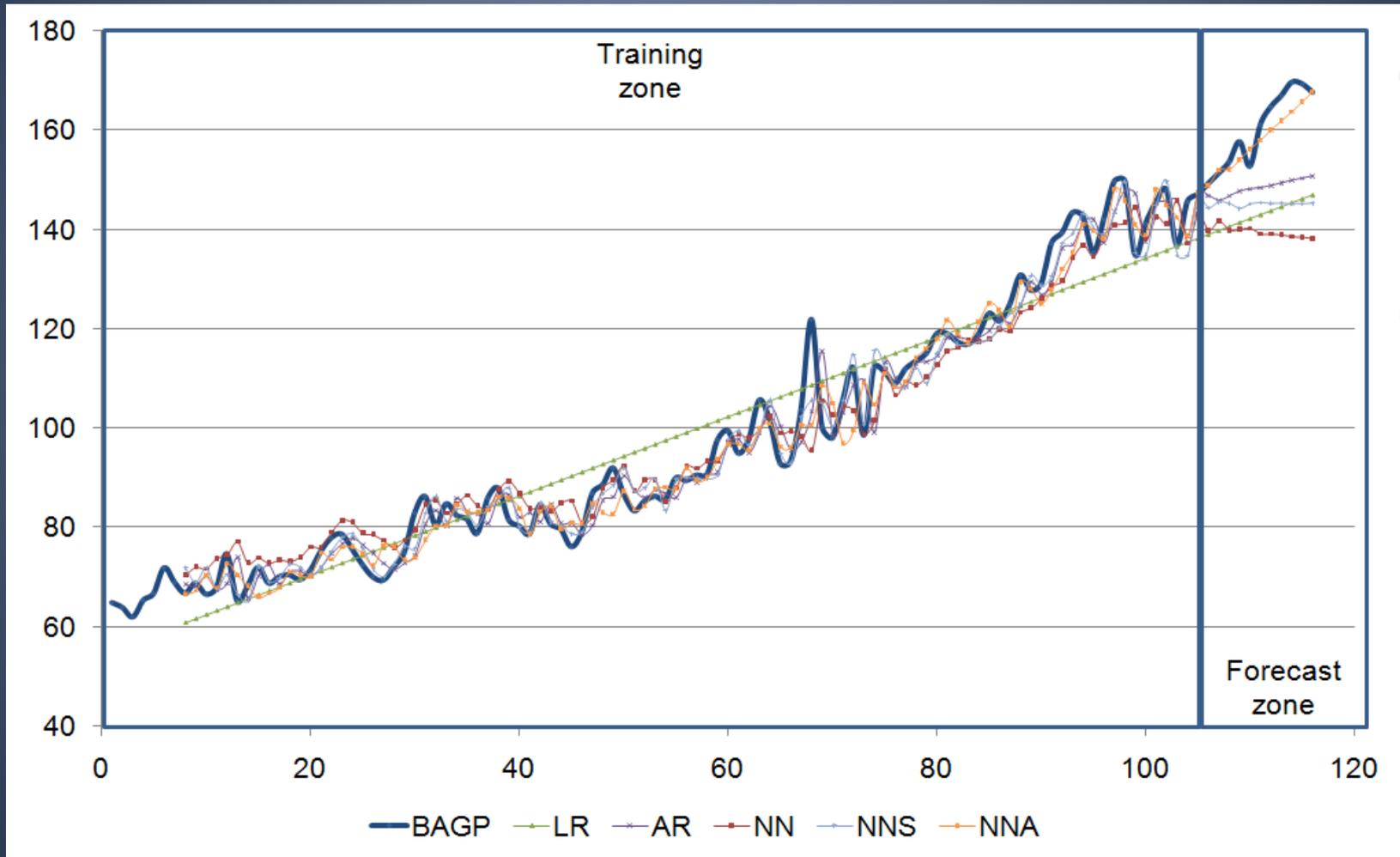
Normalização adaptativa

- Transformação
 - Conversão da série em janelas deslizantes
 - Cálculo de inércia da janela
 - Diferenciação das observações pela inércia
 - Análise da distribuição e remoção de outliers
 - Normalização
- Predição
 - Desnormalização
 - Adição da inércia

Intuição

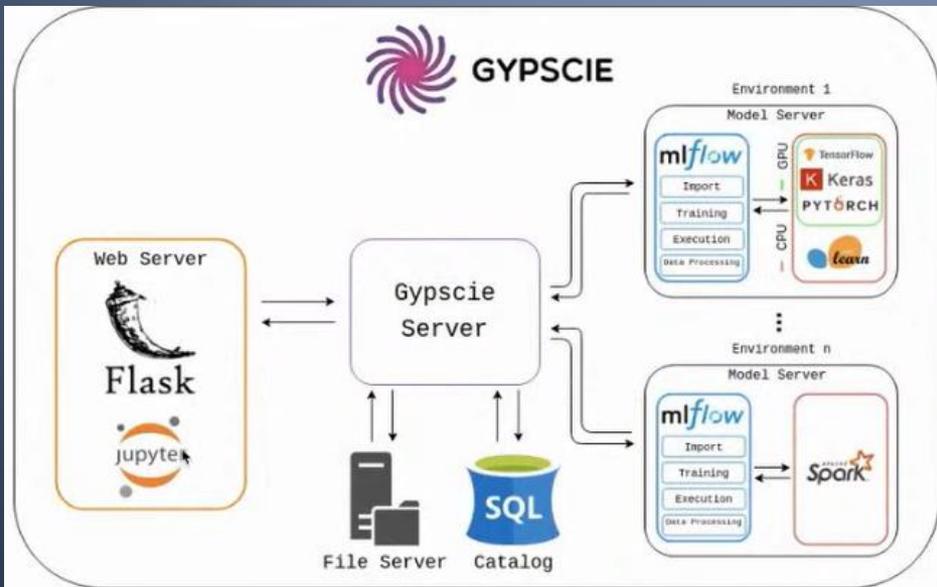


Predição de séries temporais usando aprendizado de máquina



Gerência de Modelos

- Gerência de Modelos
 - Construção
 - Produção
 - Reutilização
 - Transferência de conhecimento
 - Seleção dinâmica
 - Distribuição
 - Espaço-tempo



Ciclo de Seminários Online Instituto de Computação
Gypscie: Um Ambiente para Gerência de Dados e Modelos de Aprendizado de Máquina
Fabio Porto
LNCC
TRANSMISSÃO AO VIVO
15 de Outubro
17:00
DEXL DATA EXTREME LAB
Laboratório Nacional de Computação Científica
YouTube http://tiny.cc/ucff_youtube <https://fb.me/Informeic>
#InstitutoDeComputação #UFFContraCoronaVirus #FiqueEmCasa

Desafios

- Aprendizado semi-supervisionado /não-supervisionado
- Estruturas complexas/heterogêneas
- Streaming
- Drifts transientes
- Classes desbalanceadas
- Big Data e Small Data
- Frameworks teóricos