

Aprendizado de Máquina Supervisionado para Séries Temporais na Área da Saúde



Link para esta
apresentação

bit.ly/slides-sbcas2023



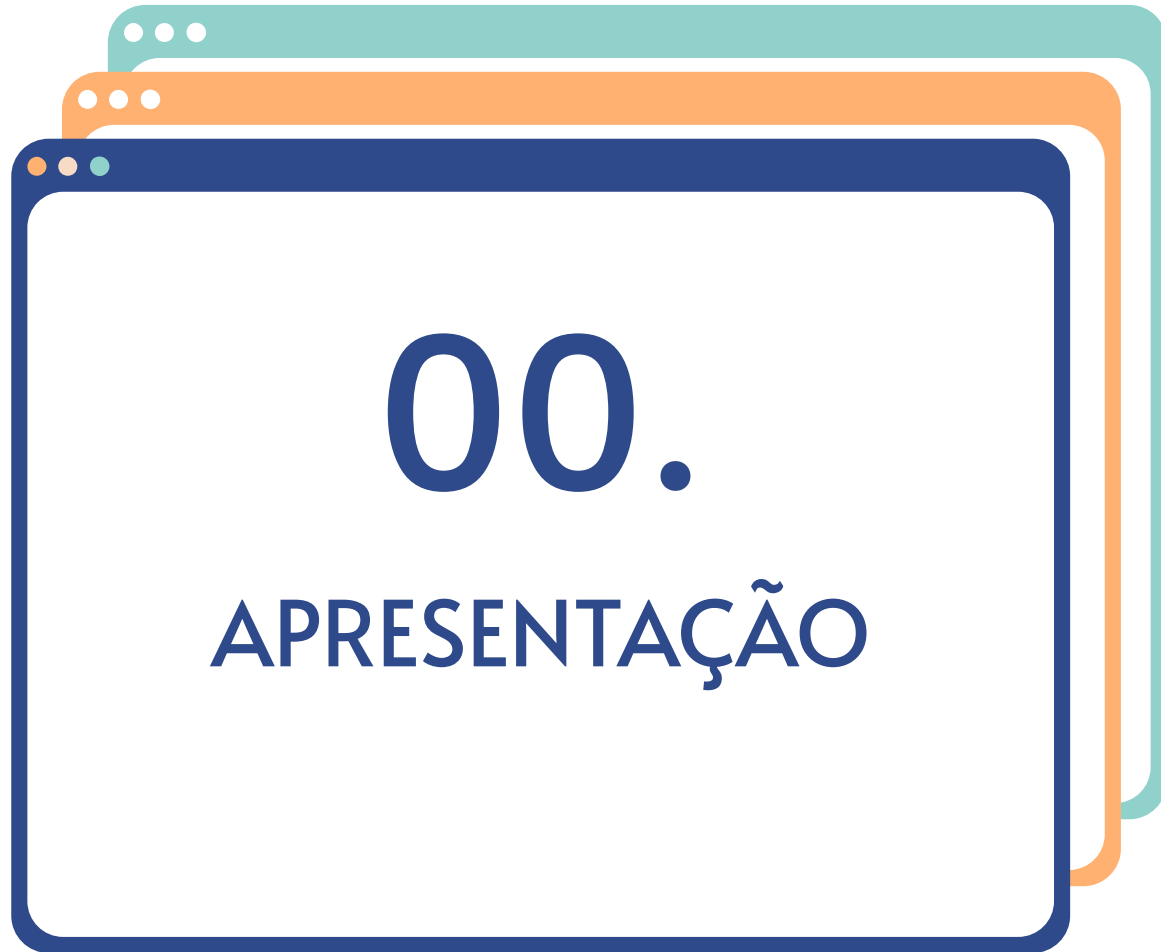
Link para o Jupyter
notebook com as
demos

bit.ly/colab-sbcas2023

Aprendizado de Máquina Supervisionado para Séries Temporais na Área da Saúde

Diego F. Silva, Guilherme G. Arcencio, José Gilberto B. M. Júnior, Vinícius M. A. de Souza, **Yuri G. A. da Silva**





QUEM SOMOS NÓS



Diego Furtado Silva
Professor @ ICMC-USP



Yuri Gabriel A. da Silva
Pesquisador DD @ ICMC-USP

Quem somos nós

Demais autores:

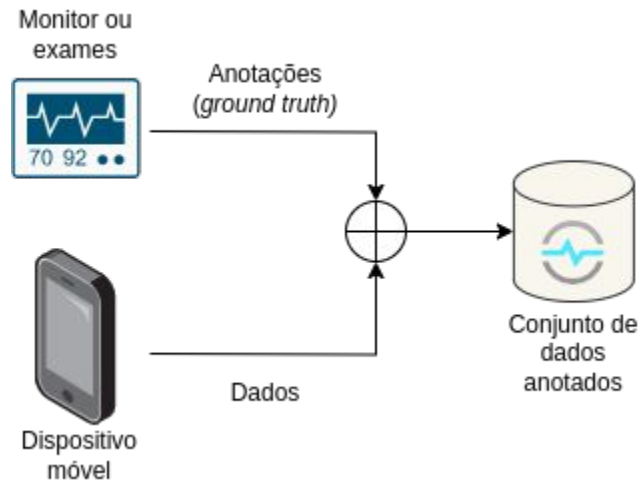
- Guilherme G. Arcencio (DC-UFSCar)
- José Gilberto B. M. Júnior (ICMC-USP)
- Vinícius M. A. de Souza (PPG1a-PUCPR)

Outros colaboradores: Audrey Silva (DFisio-UFSCar), Thiago Mazzu (DMed-UFSCar), colaboradores no exterior, outros estudantes

Quem somos nós

Colaboradores?

Projeto **FAPESP** #2022/03176-1 - Aprendizado de Máquina para Séries Temporais em Aplicações de mHealth



Quem somos nós

Inicialmente, tentaremos estimar os parâmetros do hemograma

SÉRIE BRANCA			
RESULTADO	VALORES DE REFERÊNCIA (Masc. Acima de 16 anos)		
	%	/mm ³	/mm ³
LEUCÓCITOS		7.540	3.500 a 10.500
Neutrófilos	: 52,5	3.960	1.700 a 7.000
Eosinófilos	: 3,6	270	50 a 500
Basófilos	: 0,4	30	0 a 300
Linfócitos	: 36,2	2.730	900 a 2.900
Monócitos	: 7,3	550	300 a 900

CARACTERES MORFOLÓGICOS:

não foram observados caracteres tóxico-degenerativos nos neutrófilos; não foram observadas atipias linfocitárias

AGENDA

01.

Introdução -
Dados na
Saúde

02.

Aprendizado
de Máquina

03.

Séries
Temporais

04.

AM para
Séries
Temporais



01.

INTRODUÇÃO

Dados na saúde

Dados na saúde

A utilização de dados em aplicações para a Saúde não é novidade

- Imagens (“estáticas” ou vídeos) são muito comuns
- Dados de internação
- Diagnóstico assistido
- Sinais vitais
- ...

Dados na saúde

A tendência é aumentar o número de aplicações, além de dar mais “poder ao paciente”

- Armazenamento
- Sensores móveis
- Telemedicina

Dados na saúde

“Digital technologies, such as mobile wireless technologies, have the potential to **revolutionize how populations interact with national health services.**”

World Health Organization. (2018). mHealth. Use of appropriate digital technologies for public health: report by Director-General. 71st World Health Assembly provisional agenda item, 12, A71.

Dados na saúde

Alguns exemplos (referências no texto)

- Monitoramento de pacientes
 - Doença de Parkinson
 - Resposta à medicação
 - Prevenção a ataques cardíacos
 - etc

Dados na saúde

A pergunta do curso é “o que podemos fazer com esses dados?”.

Como é impossível cobrir todos os tipos de dados, vamos focar em um deles: as séries temporais!



02.

SÉRIES TEMPORAIS

Definição e exemplos

Definição

Grosso modo:

- dados observados/coletados no tempo
- Mas não pode ser só isso!



Definição

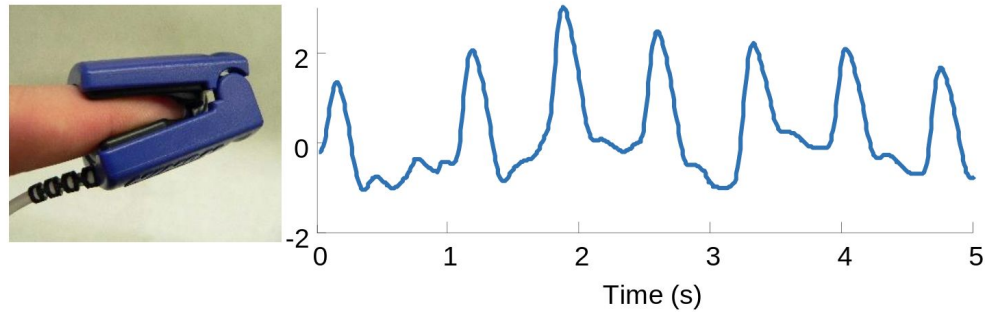
Uma série temporal é um **conjunto de valores ordenados** S , tal que:

- $S = (s_1, s_2, \dots, s_n)$
- $s_t \in \mathbb{R}, \quad \forall t \in [1, \dots, n]$

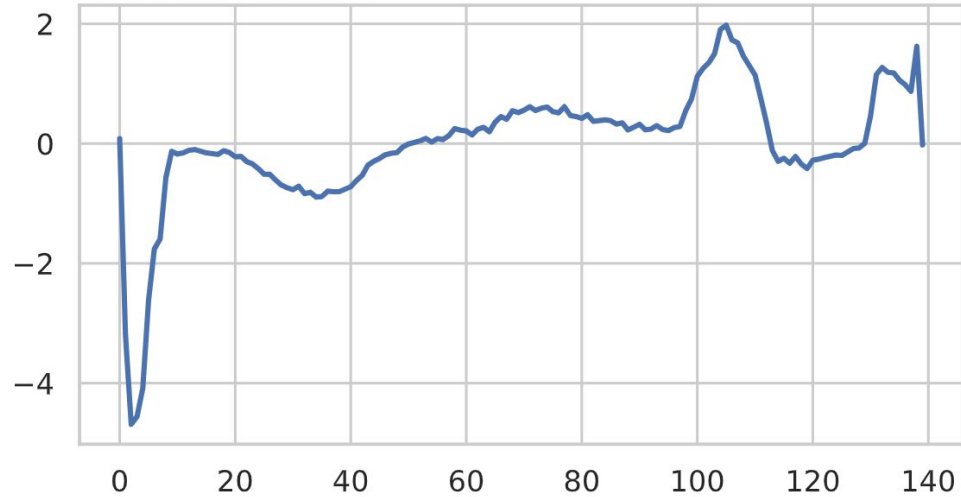
Ou seja, não precisa haver “variação no tempo”

p.s. Geralmente, considera-se espaçamento uniforme

Definição



Definição



Definição

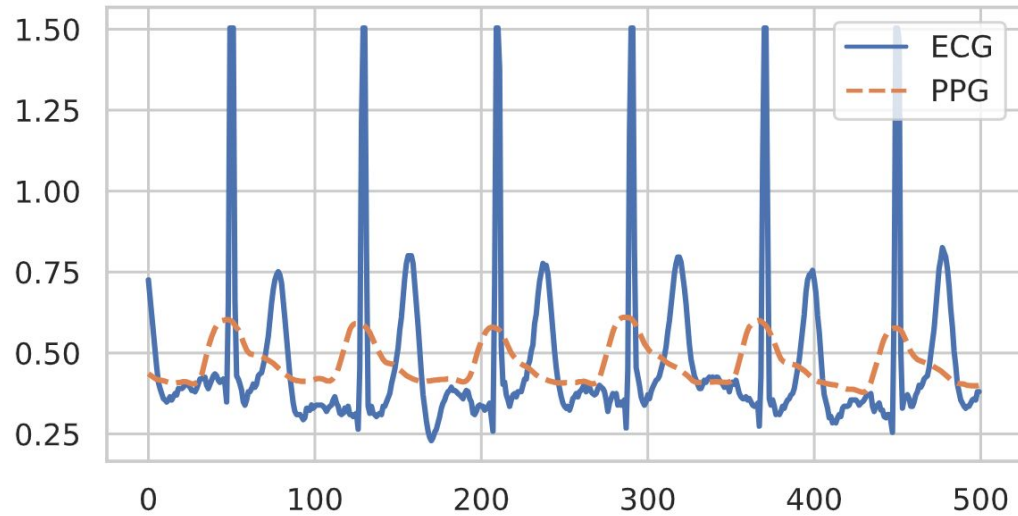
Essas eram séries temporais univariadas. Generalizando:

Uma série temporal é um conjunto de valores ordenados S , tal que:

- $S = (s_1, s_2, \dots, s_n)$
- $s_t \in \mathbb{R}^d, \quad \forall t \in [1, \dots, n]$

Univariada: $d = 1$

Definição



Definição



<https://archive.physionet.org/physiobank/database/mghdb/>

Definição

Uma subsequência \hat{S} é um subconjunto contíguo de observações da série temporal S :

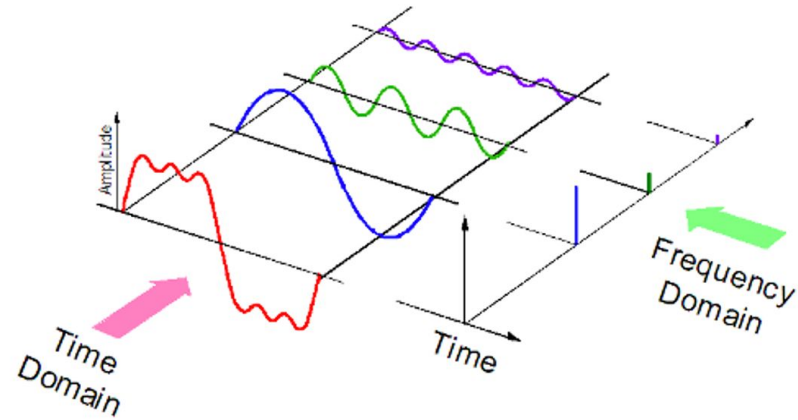
- $\hat{S} = (s_i, s_{i+l}, \dots, s_{i+m-l})$
- $s_i \in S, \quad \forall i \in [1, \dots, n-m]$

Domínio da Frequência

Uma série temporal pode ser decomposta por suas **componentes de frequência**

- Nesse **domínio**, podemos fazer extração de características, filtragem, entre outros

Fourier - noção geral



Obs:

Assume estacionariedade (já falo sobre).

Caso contrário, tempo-frequência

Pré-processamento

Analytics

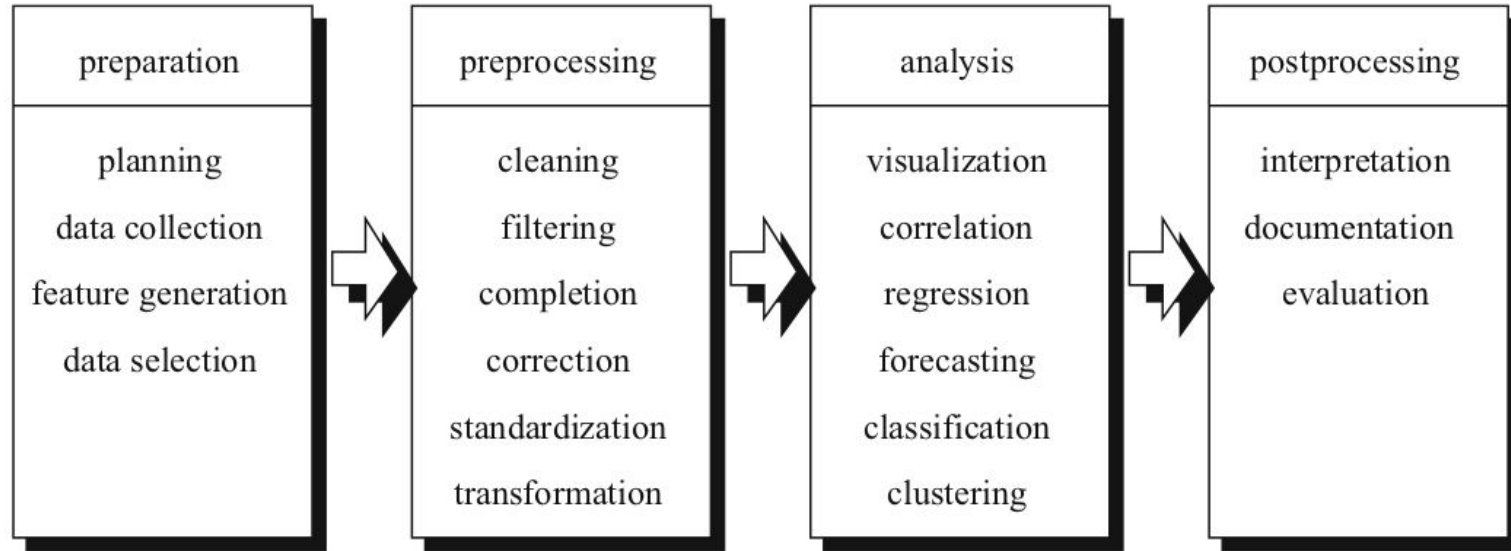
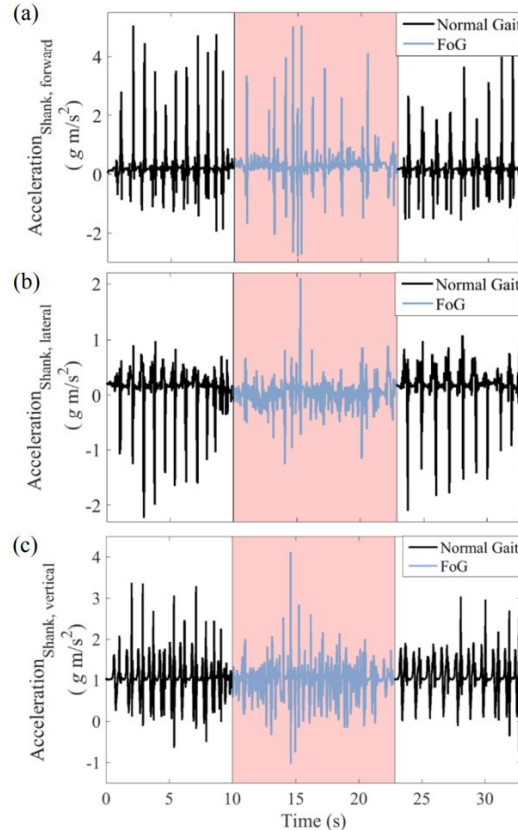


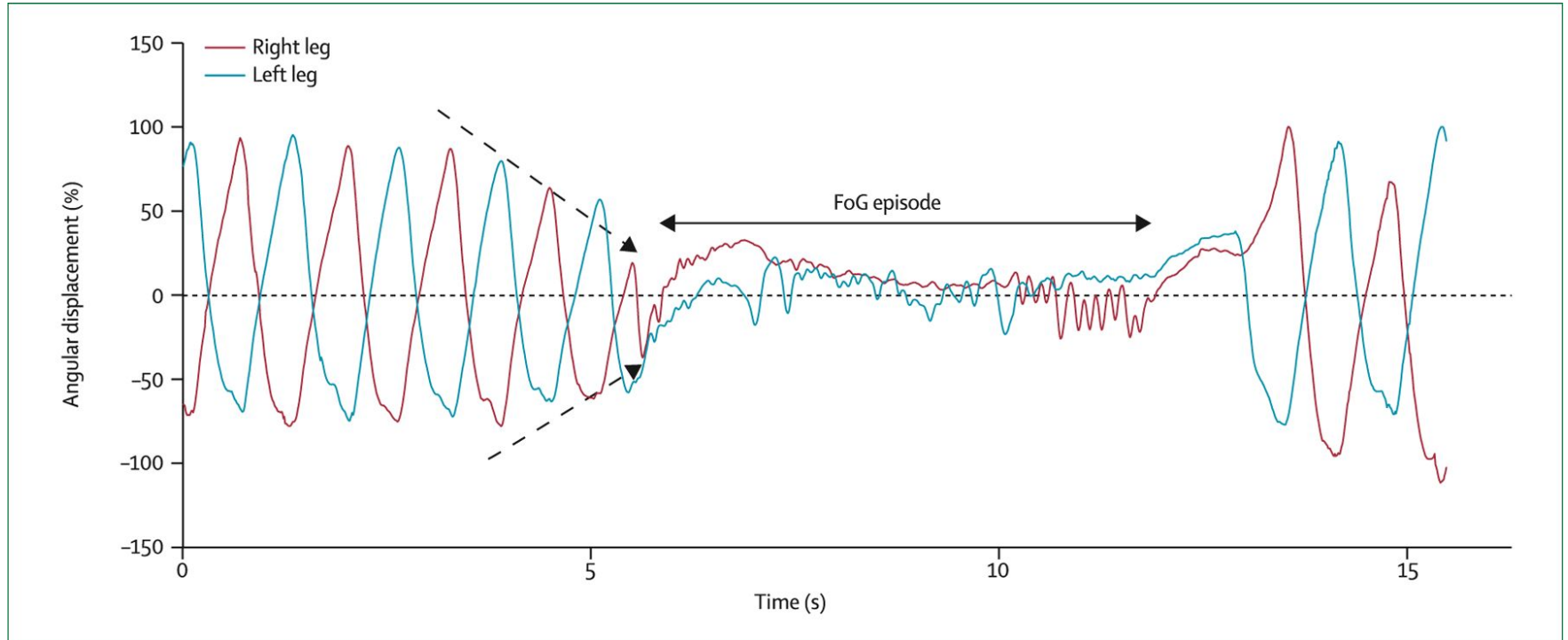
Fig. 1.1 Phases of data analysis projects

Exemplo da preparação



ARAMI, Arash et al. Prediction of gait freezing in parkinsonian patients: A binary classification augmented with time series prediction. IEEE Transactions on Neural Systems and Rehabilitation Engineering, v. 27, n. 9, p. 1909-1919, 2019.

Exemplo da preparação



Analytics

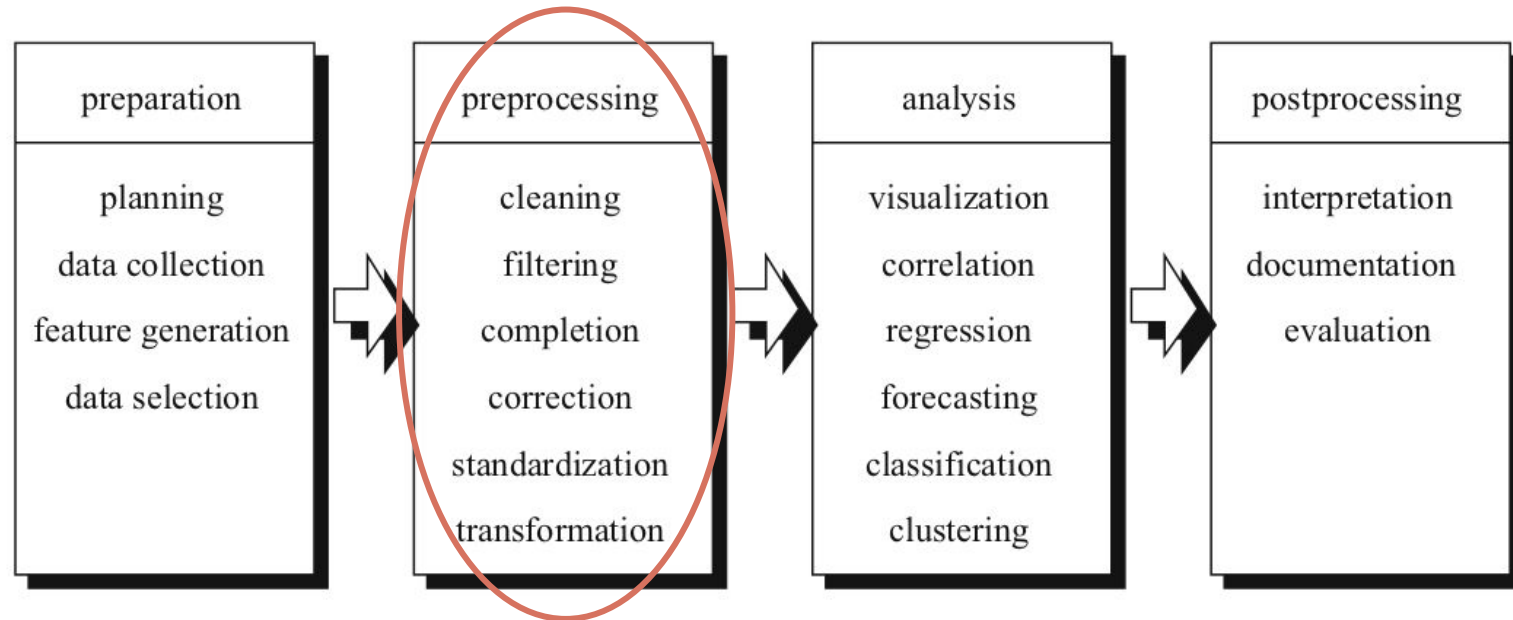
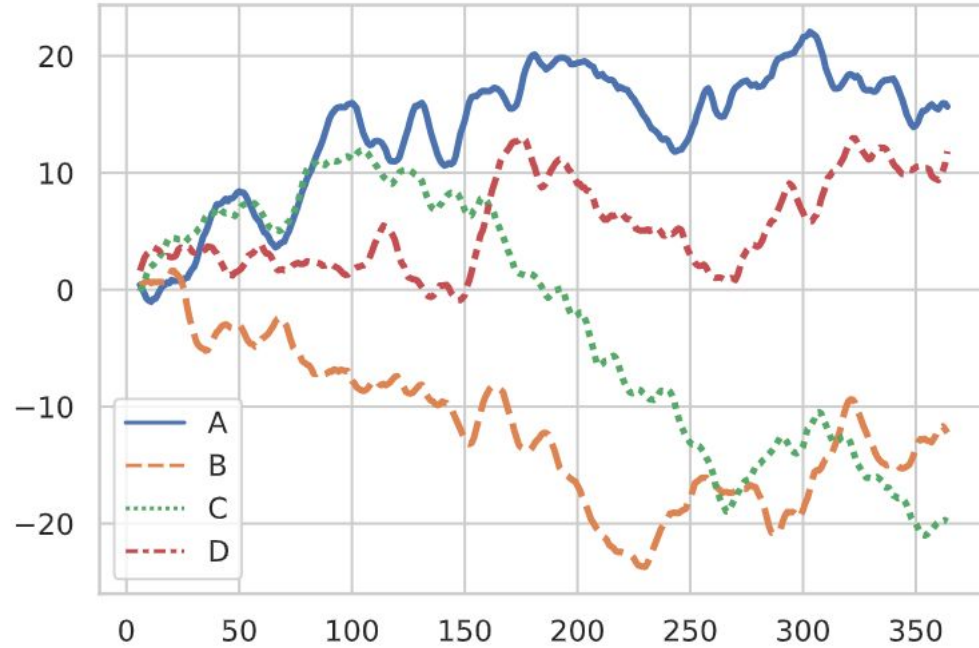
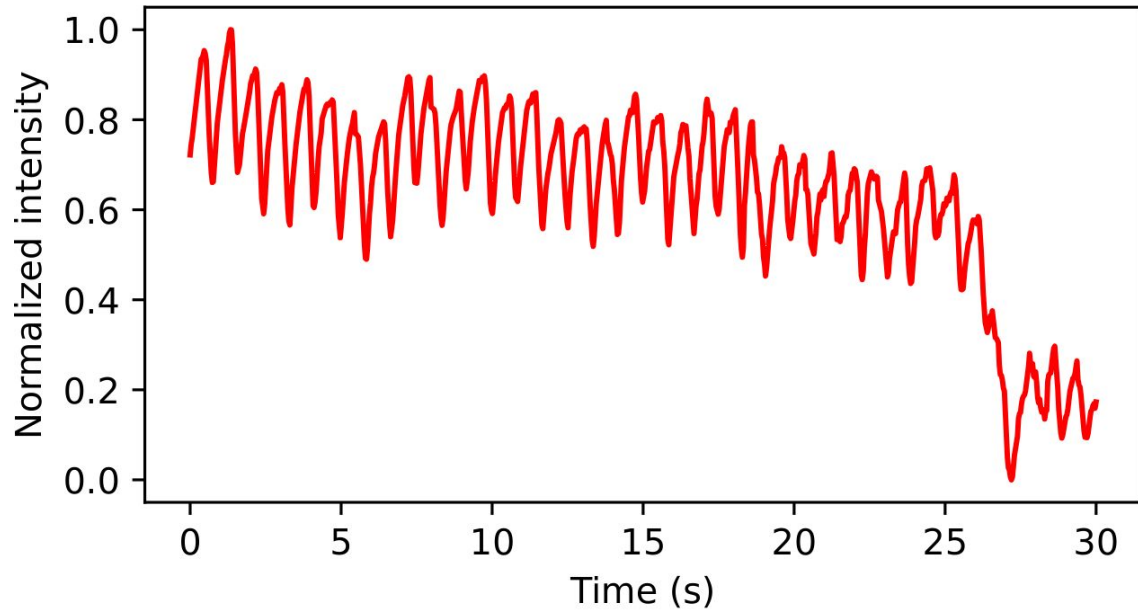


Fig. 1.1 Phases of data analysis projects

Componentes



Componentes



Componentes

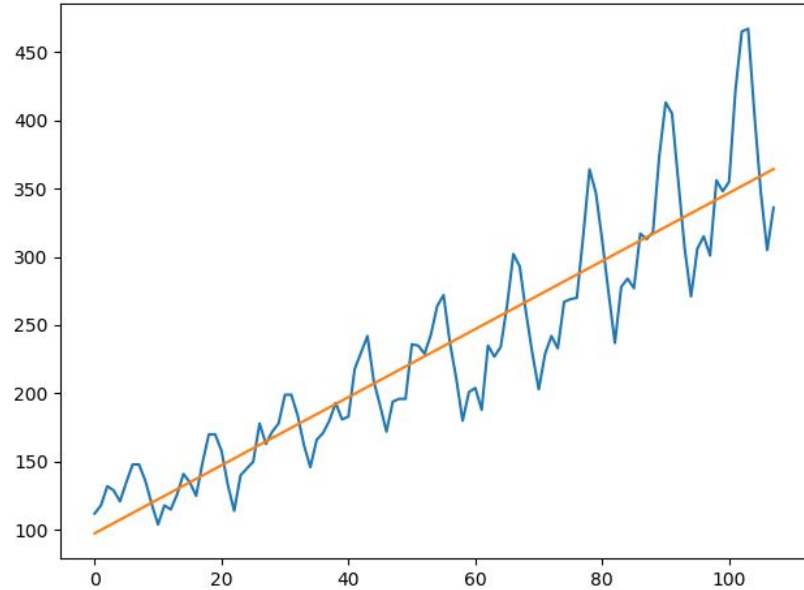
- Tendência
- Sazonalidade
- Ciclicidade
- Residual ou irregularidade (componente aleatoriedade)

Se a tarefa for o *forecasting*, essa decomposição é ainda mais importante, além de ser interessante sabermos o conceito de estacionariedade:

- Quando uma subsequência $(x_t, x_{t+p}, \dots, x_{t+m})$ tem a mesma distribuição conjunta de $(x_{t+j}, x_{t+j+p}, \dots, x_{t+j+m})$, para qualquer j inteiro

Componentes

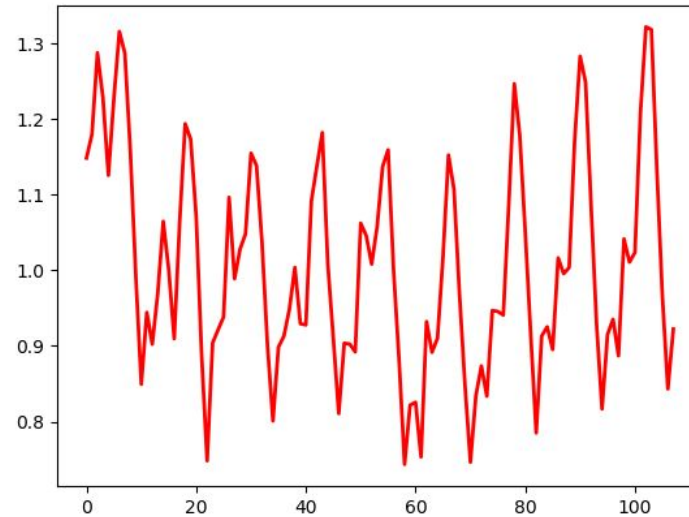
- **Tendência**
- **Sazonalidade**
- Ciclicidade
- Residual



Componentes

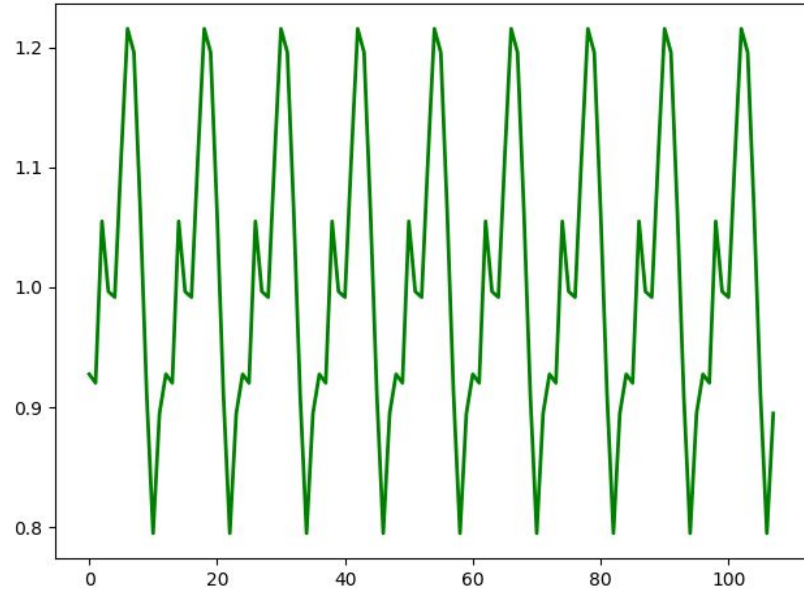
- **Tendência**
- **Sazonalidade**
- Ciclicidade
- Residual

Dados sem a tendência (X/trend)



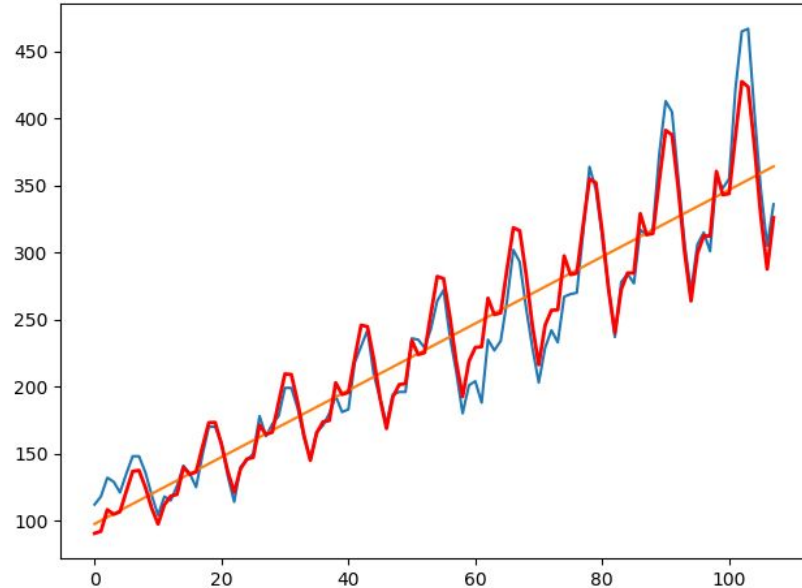
Componentes

- **Tendência**
- **Sazonalidade**
- Ciclicidade
- Residual



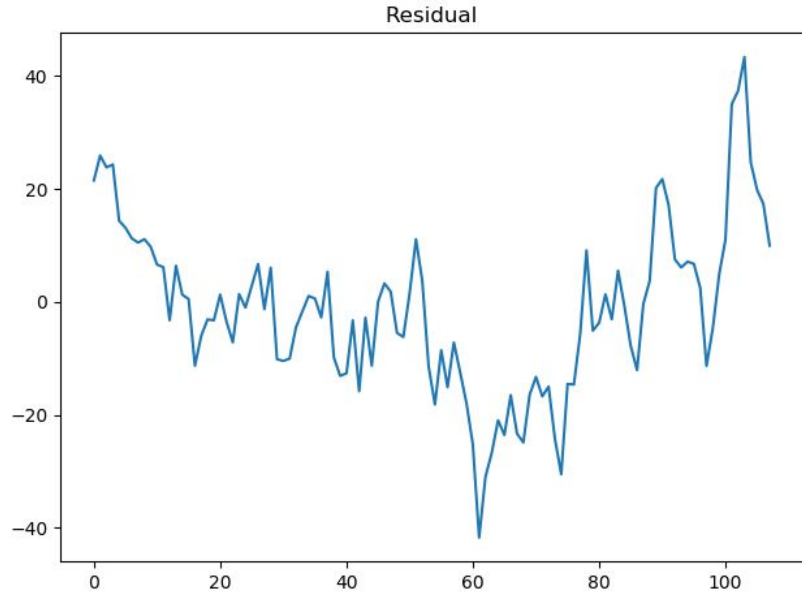
Componentes

- **Tendência** * (se multiplicativo)
- **Sazonalidade** = aproximação da série
- Ciclicidade
- Residual



Componentes

- **Tendência**
- **Sazonalidade**
- Ciclicidade
- **Residual**



Decomposição

Acabamos de **decompor** a série temporal

Porém, eu fiz isso usando regressão linear e outras ferramentas “fáceis”.

- Há ferramentas muito melhores para isso
 - No exemplo a seguir, utilizei a lib statsmodel

Decomposição

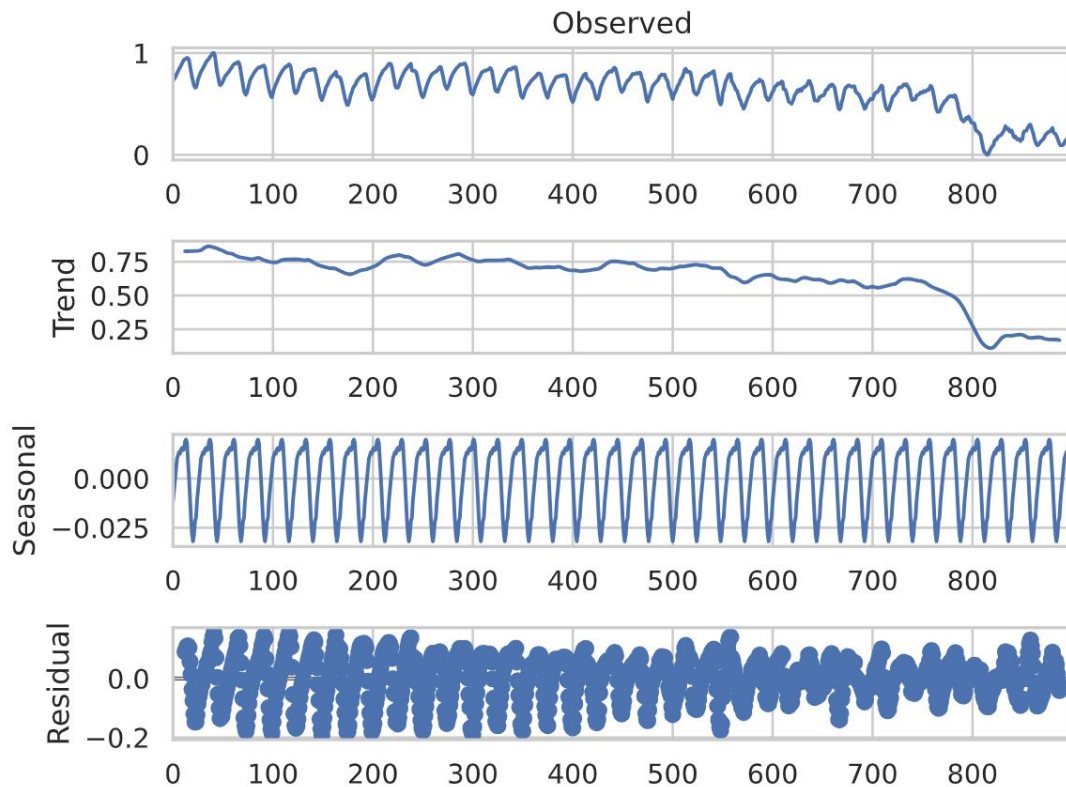
```
1 from statsmodels.tsa.seasonal import seasonal_decompose
2
3 import matplotlib.pyplot as plt
4 import numpy as np
5
6 series = np.genfromtxt('data.tsv', delimiter='\t')
7 result = seasonal_decompose(series,
8                             model='additive',
9                             period=24)
10
11 plt.figure(figsize=[6,9])
12 result.plot()
13 plt.show()
```

Códigos aqui

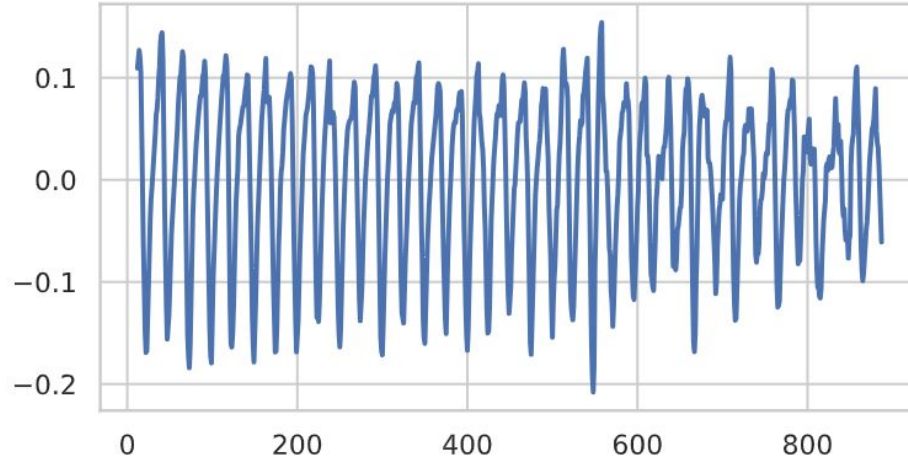


bit.ly/colab-sbcas2023

Decomposição



Detrend

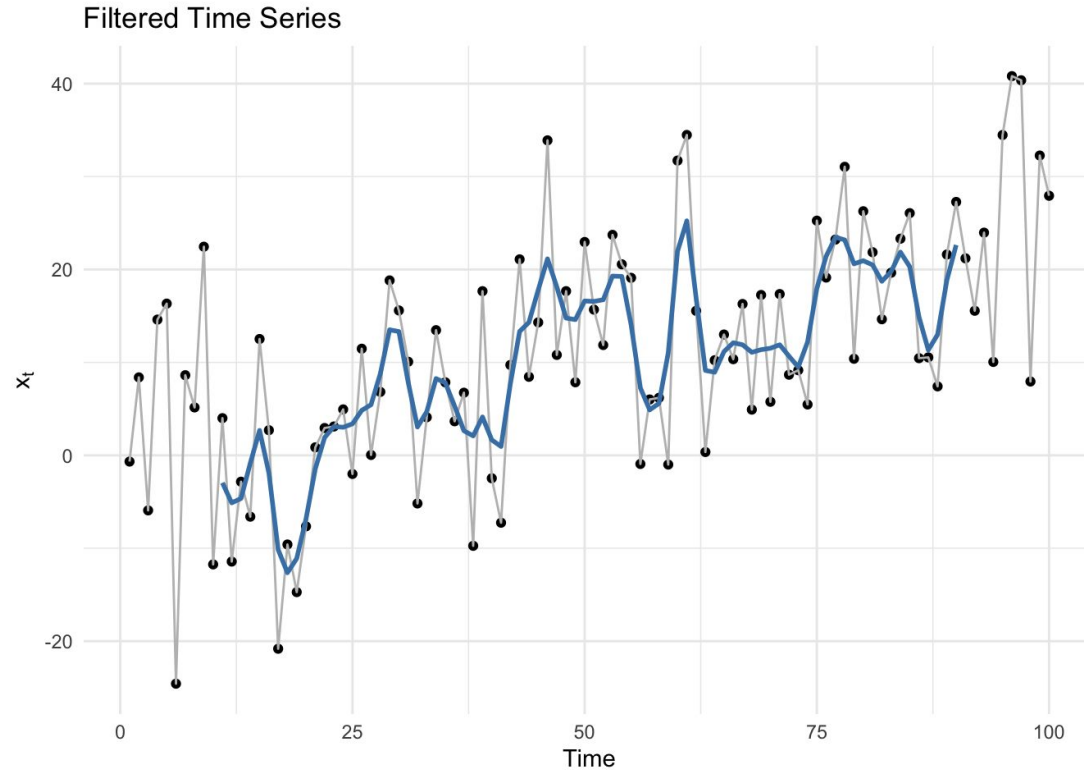


Filtragem (além da tendência)

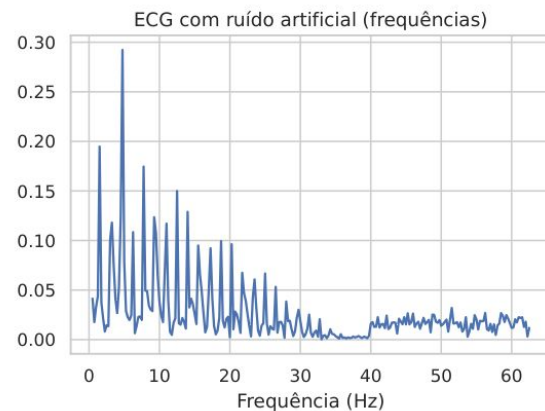
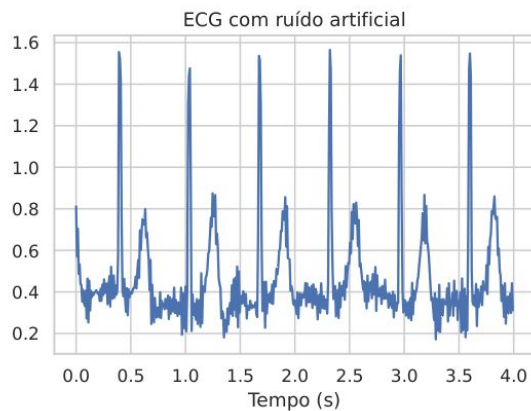
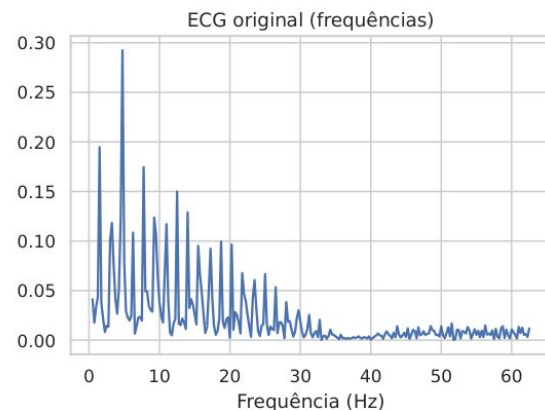
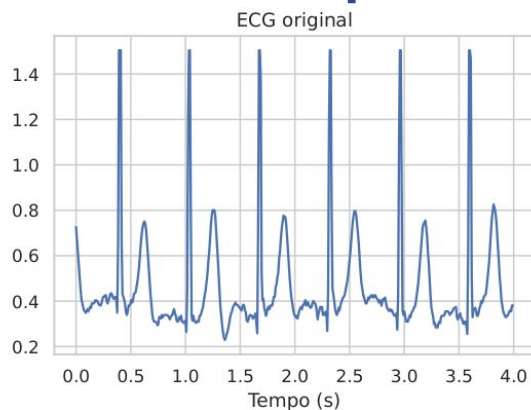
Por exemplo,

- para diminuir a “complexidade” da série temporal, utilizamos filtros de média móvel
- para suprimir picos espúrios, utilizamos filtros de mediana móvel
- para se remover ruídos em frequências altas ou baixas, utilizamos filtros de passa banda

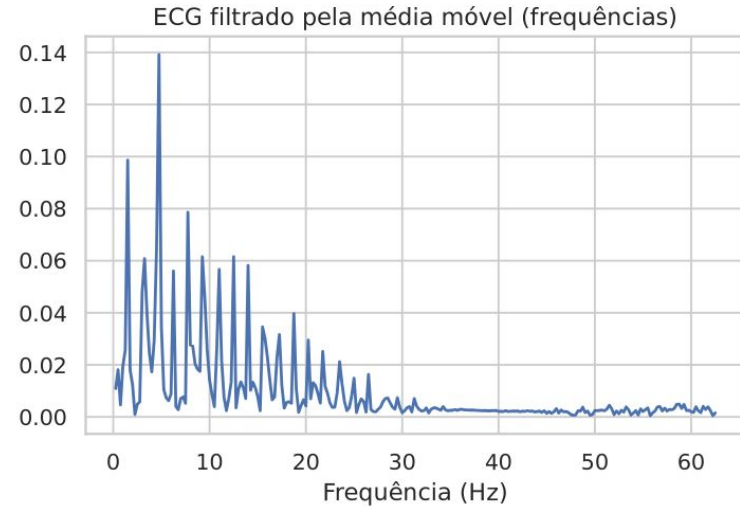
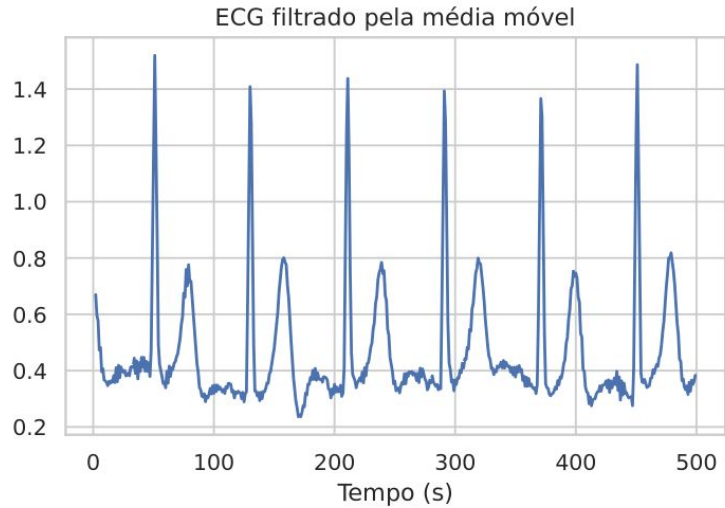
Filtragem - passa baixa



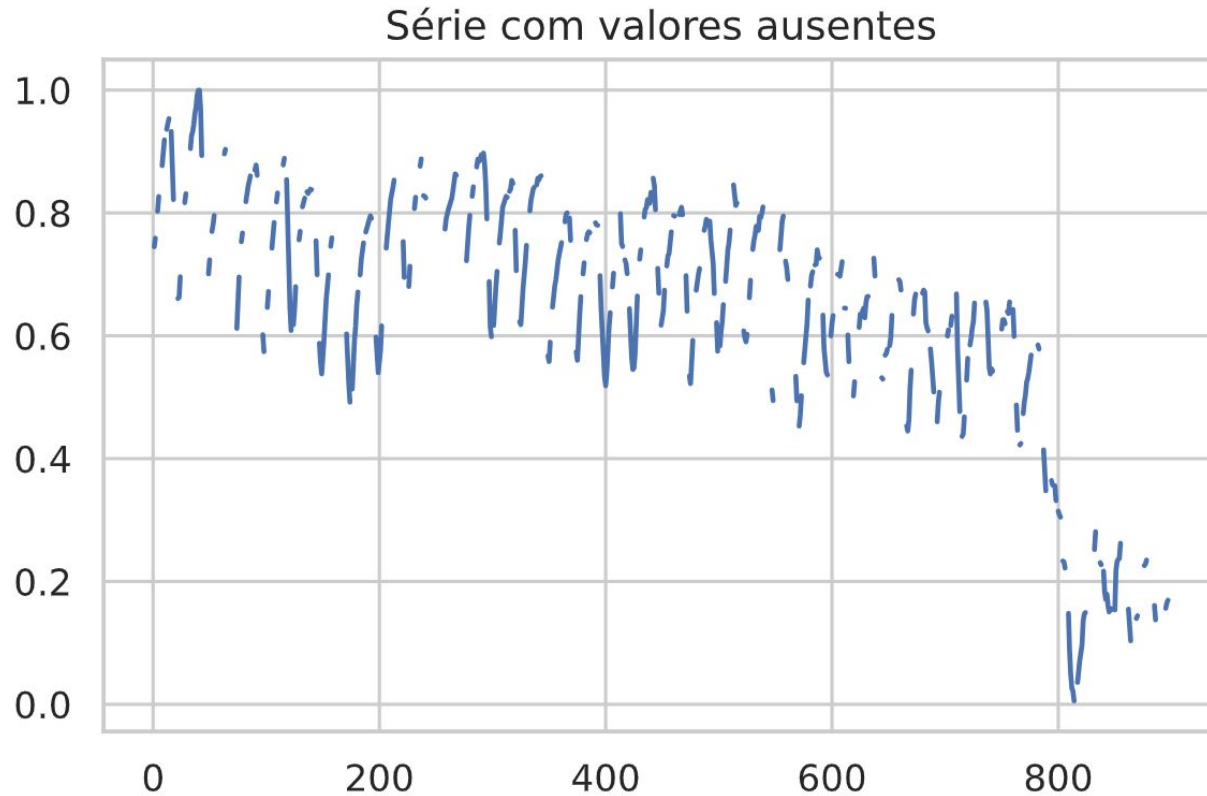
Filtragem - exemplo com média móvel



Filtragem - exemplo com média móvel

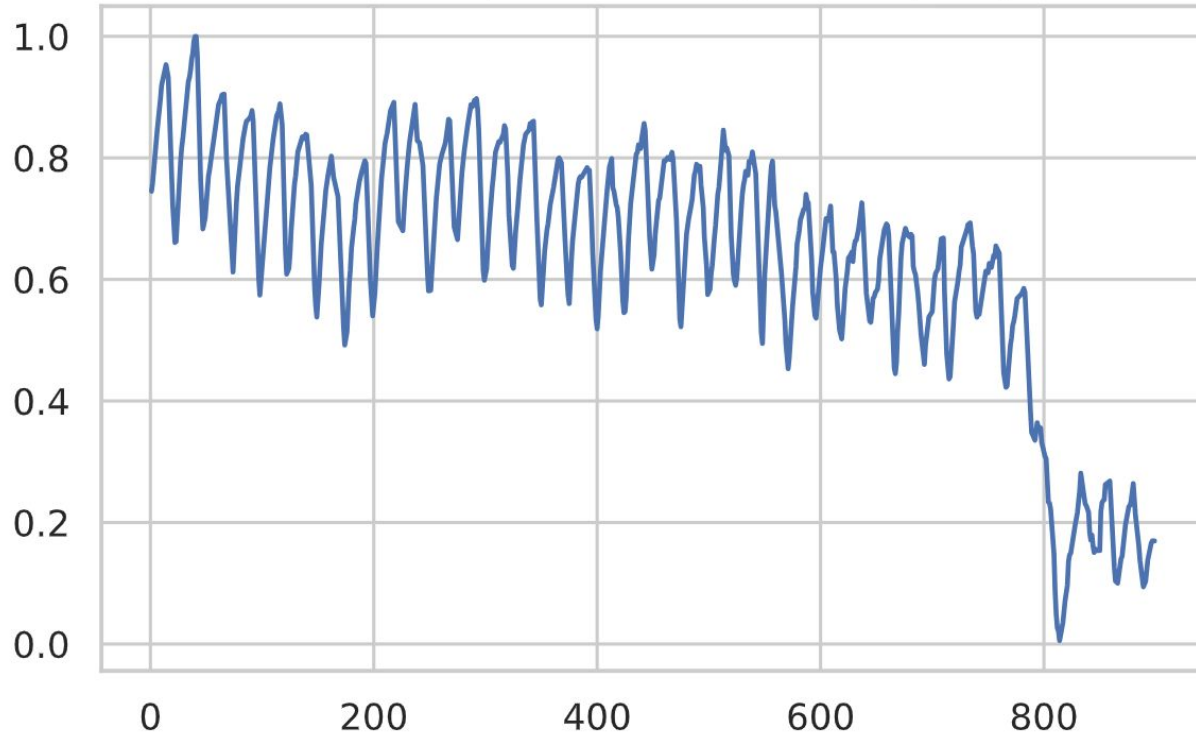


Interpolação

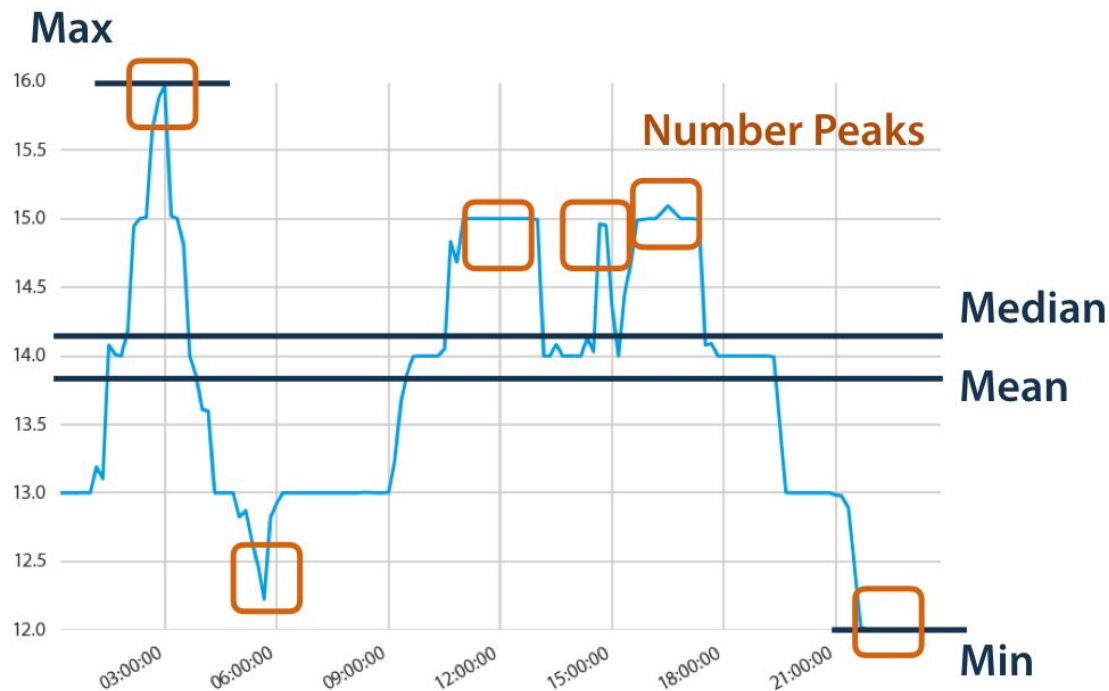


Interpolação

Série com valores preenchidos por interpolação linear

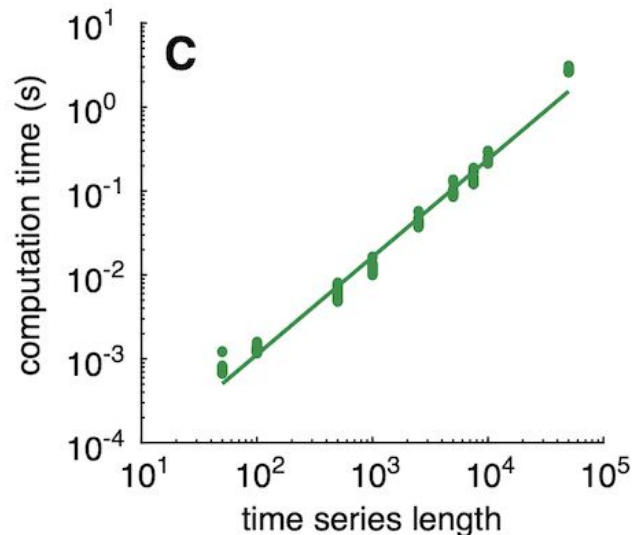
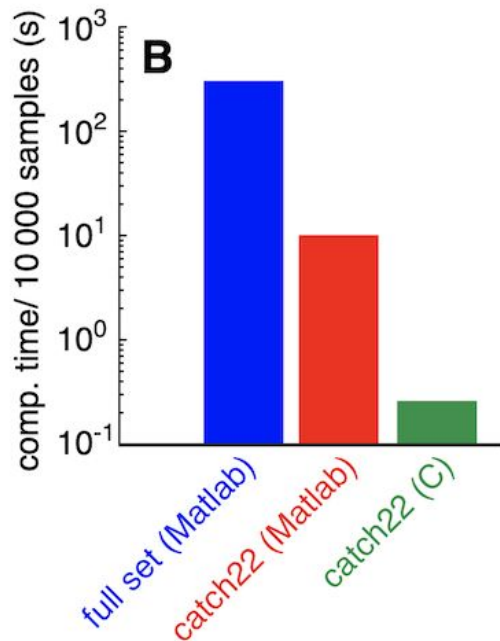
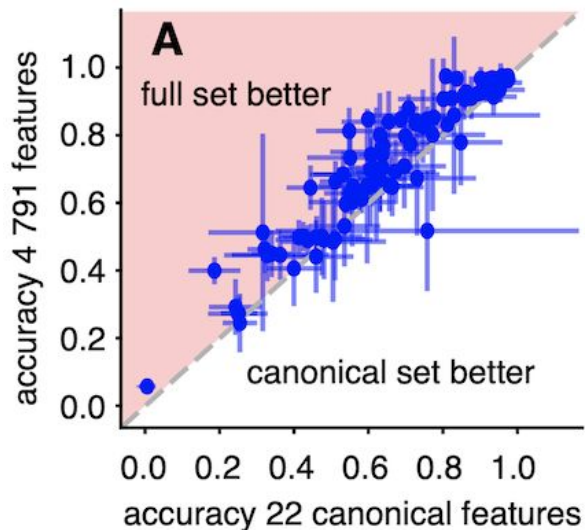


Extração de características

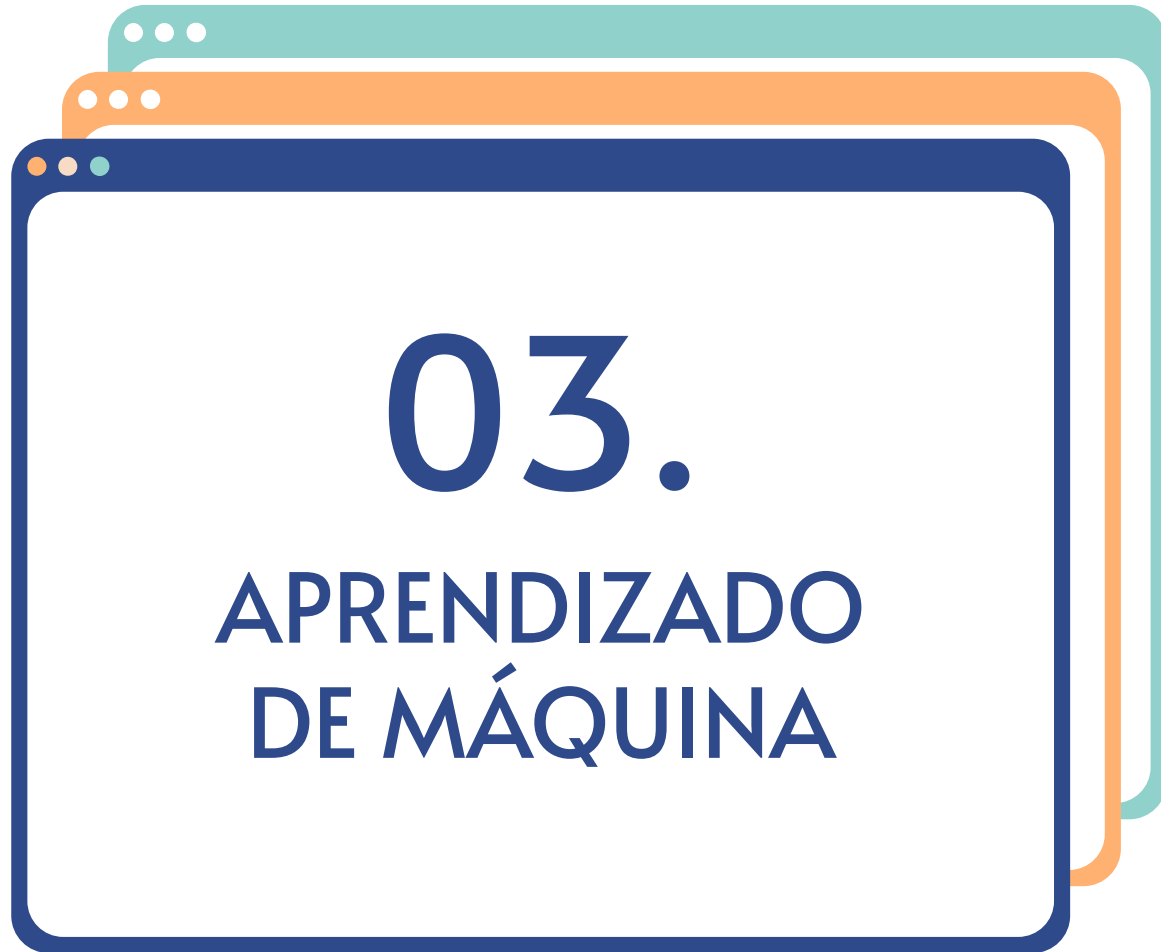


<https://tsfresh.readthedocs.io/en/latest/text/introduction.html>

Extração de características



<https://github.com/DynamicsAndNeuralSystems/catch22>



03.

APRENDIZADO
DE MÁQUINA

AM - Definições

“Machine Learning: the field of study that gives computer the **ability to learn** without being **explicitly programmed.**”

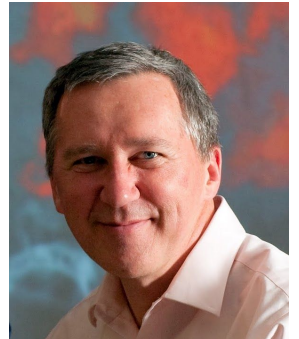
Arthur Samuel, 1959



AM - Definições

“A computer programming is said to learn from **experience E** with respect to some **task T** and some **performance** measure **P** if its performance on **T**, measured by **P**, improves with **E**.”

Tom Mitchell, 1998



AM - Exemplo

Queremos um sistema para nos auxiliar a detectar pneumonia e, de preferência, nos diga qual o tipo.

Como o Aprendizado de Máquina pode ser útil nisso?
O que são T, E e P (da premissa do Mitchell), nesse caso?



KERMANY, Daniel S. et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *cell*, v. 172, n. 5, p. 1122-1131. e9, 2018.

AM - Definição mais formal

O que compõe o aprendizado no exemplo das imagens de raio-x?

Entrada: x

Saída: y

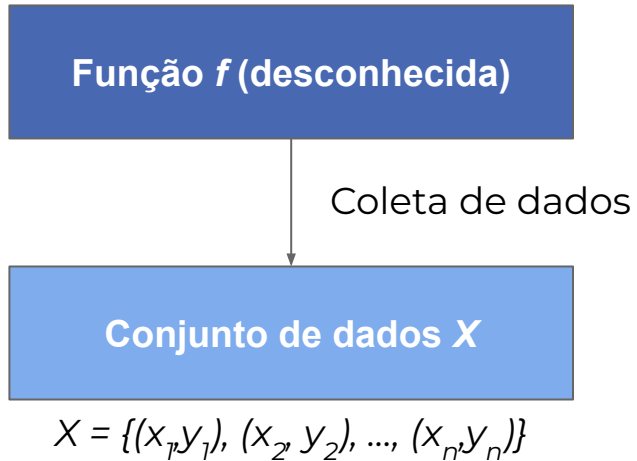
Função: $f: X \rightarrow Y$

AM - Definição mais formal

Função f (desconhecida)

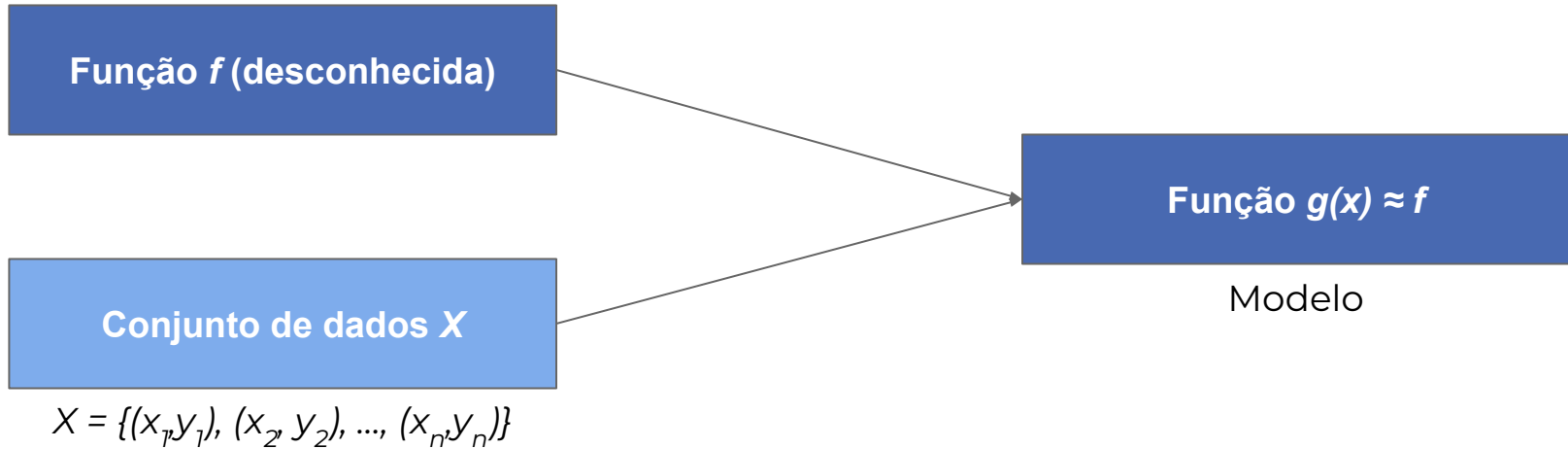
Como estimar?

AM - Definição mais formal

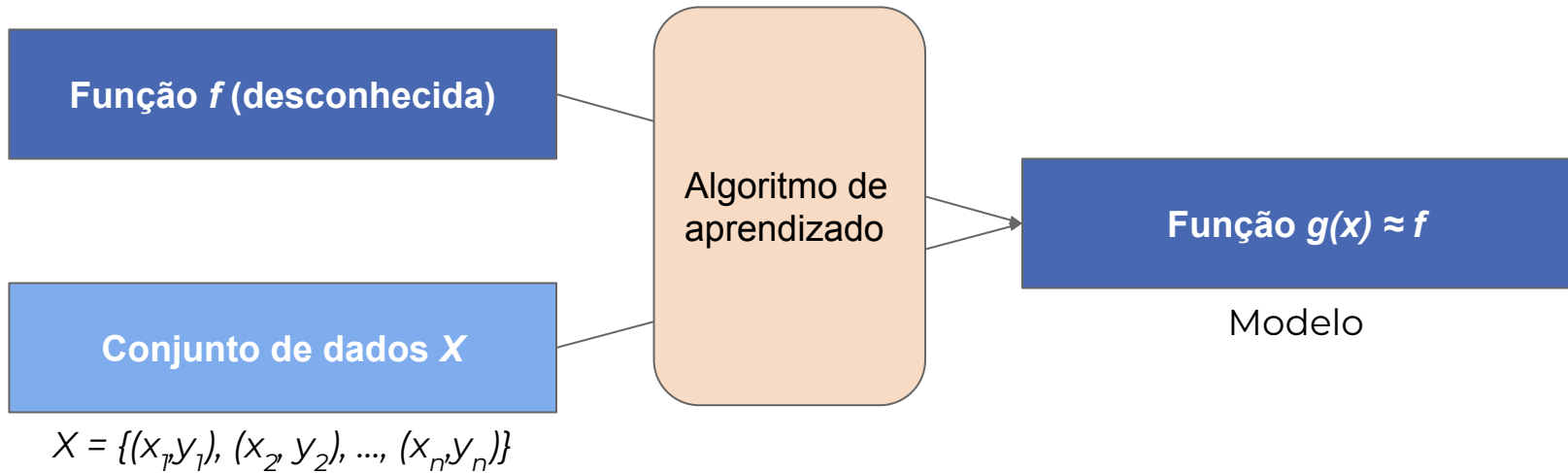


Obs: AM supervisionado

AM - Definição mais formal

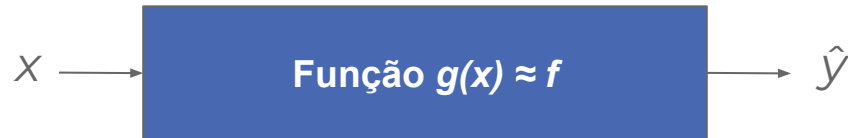


AM - Definição mais formal



AM - Modelo

No caso do AM supervisionado, o **modelo** aprendido/induzido tem por objetivo estimar um valor de atributo-alvo (\hat{y}) para uma dada entrada (x) que seria a saída da função objetivo (f) para a observação de um fenômeno que foi representado por x .



AM - Tarefas

No caso do objetivo ser prever um valor de saída (modelos preditivos), podemos ter diferentes tarefas. As duas mais comuns e que serão exploradas neste curso são:

- Classificação (atributo-alvo categórico - rótulo)
- Regressão (atributo-alvo numérico)



AM para séries

Tentem responder: o que muda quando x é uma série temporal?

AM para séries

As tarefas que abordaremos são a **classificação** e a **regressão extrínseca** (não confundir com *forecasting*)

AM para séries

Categorias de algoritmos:

- baseados em distância
- baseados em intervalos
- baseados em dicionários
- aprendizado profundo
- comitês

Onde achar tudo isso? <https://www.aeon-toolkit.org/en/latest/>

AM para séries

Observações:

- Classificação -> regressão
- Univariada -> multivariada

Distância/similaridade

Algoritmos baseados em distância

Busca encontrar o vizinho mais próximo de uma série temporal a ser classificada em uma base de séries rotuladas (1-NN)

- Equivalente para regressão
- Extensões para k-NN
- Uma das maiores vantagens: **interpretabilidade**

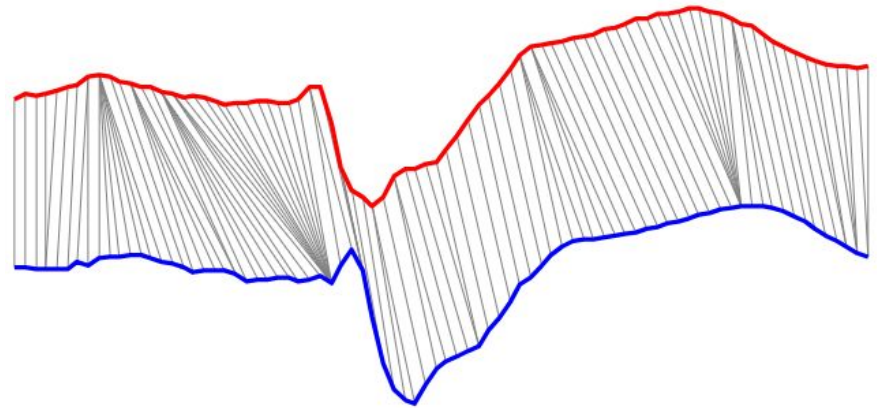
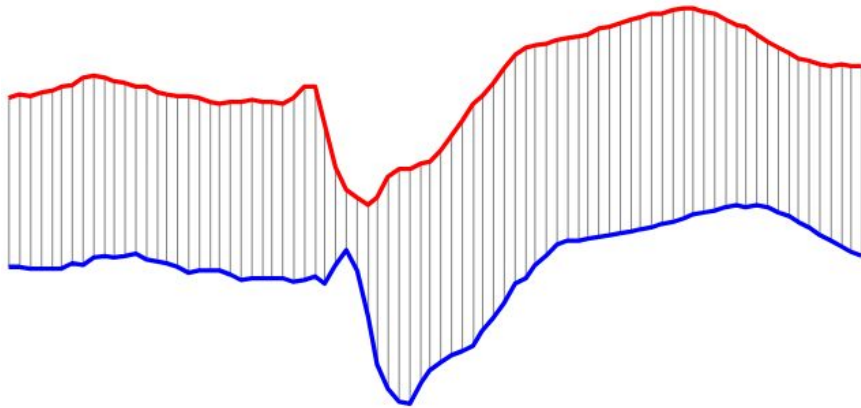
Algoritmos baseados em distância

Uma das maiores vantagens: interpretabilidade (**será?**)

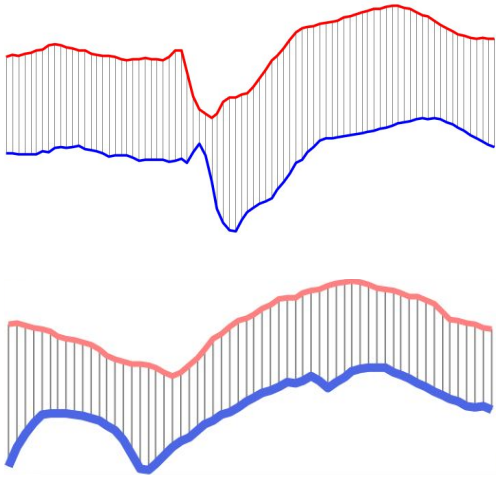


Algoritmos baseados em distância

Uma das maiores vantagens: interpretabilidade



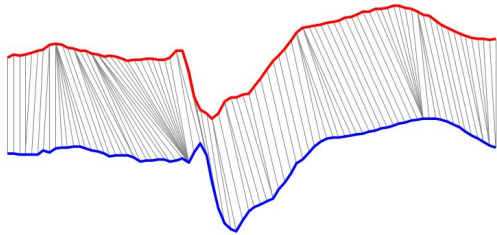
Distância Euclidiana



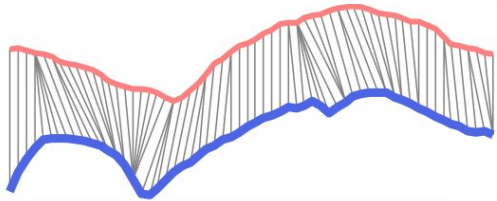
Alinhamento **linear** entre séries temporais x e y de mesmo comprimento n

$$ed(x, y) = \sqrt{\sum_{k=1}^n (x_k - y_k)^2}$$

DTW



Alinhamento **não-linear** entre séries temporais x e y , potencialmente de comprimentos diferentes

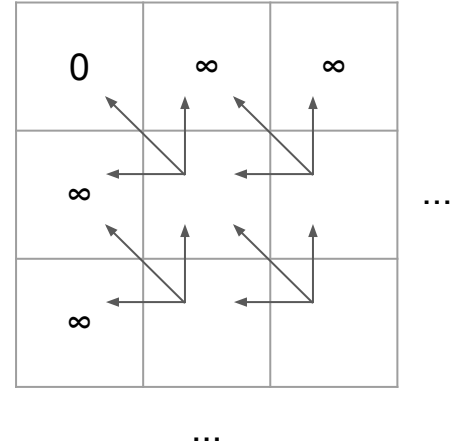


$$dtw(i, j) = c(x_i, y_j) + \min \begin{cases} dtw(i-1, j) \\ dtw(i, j-1) \\ dtw(i-1, j-1) \end{cases}$$

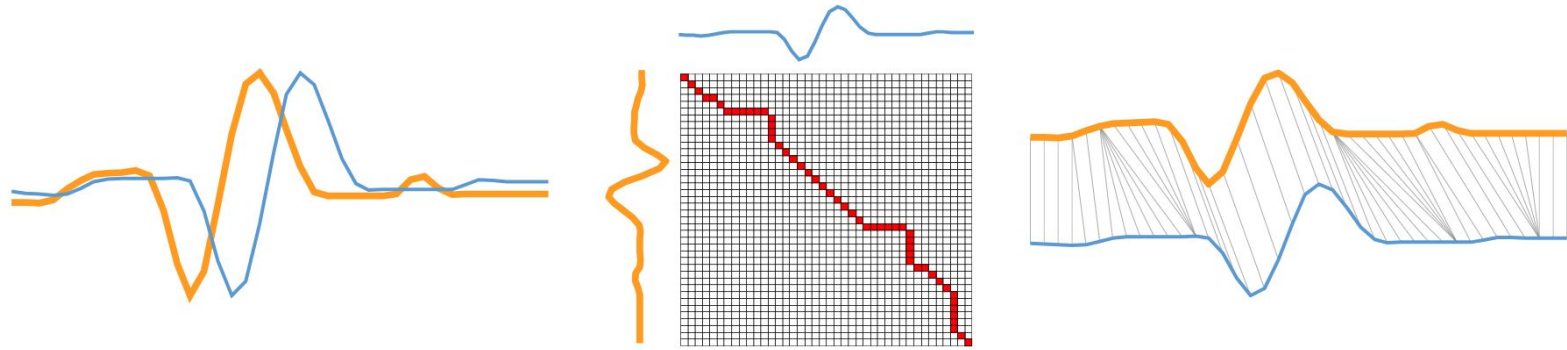
DTW

$$dtw(i, j) = c(x_i, y_j) + \min \begin{cases} dtw(i-1, j) \\ dtw(i, j-1) \\ dtw(i-1, j-1) \end{cases}$$

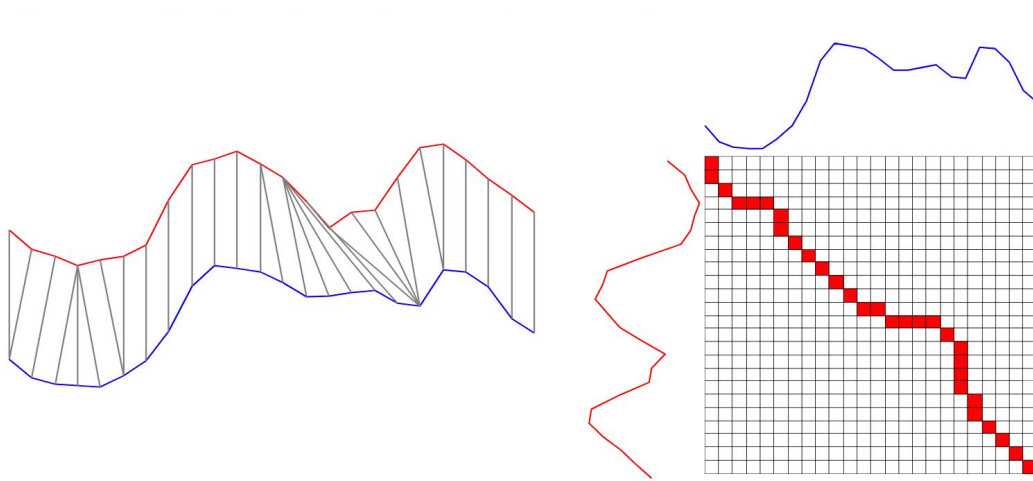
$(x_i - y_j)^2$ ↑



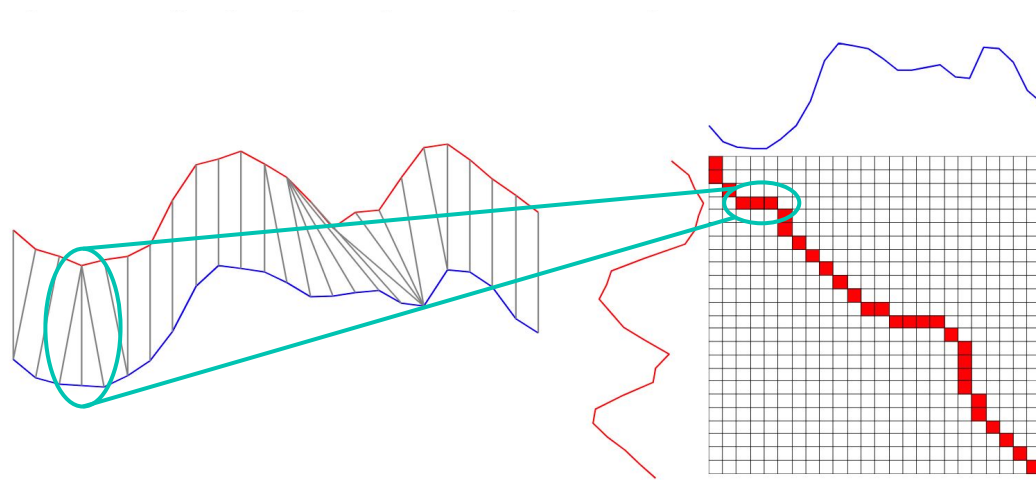
DTW



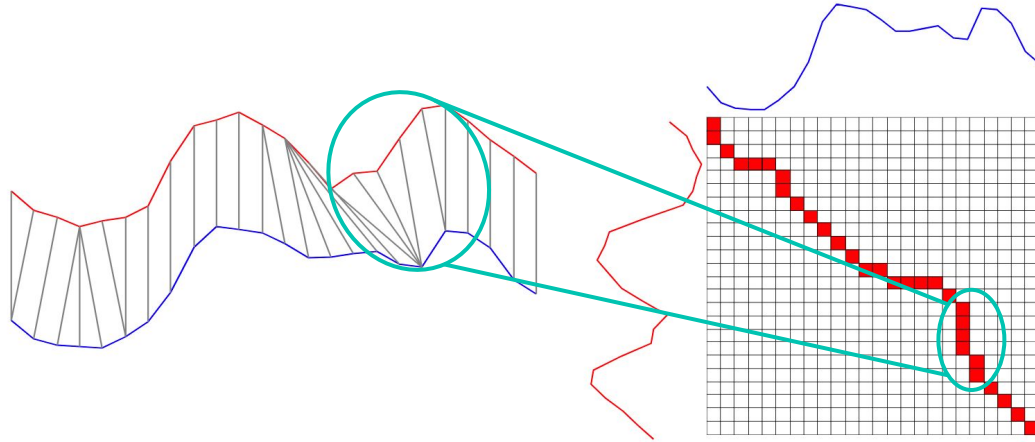
DTW



DTW



DTW



Intervalos

Algoritmos baseados em intervalos

Podemos imaginar um algoritmo que extraia características das séries e utilize qualquer algoritmo “clássico” de AM

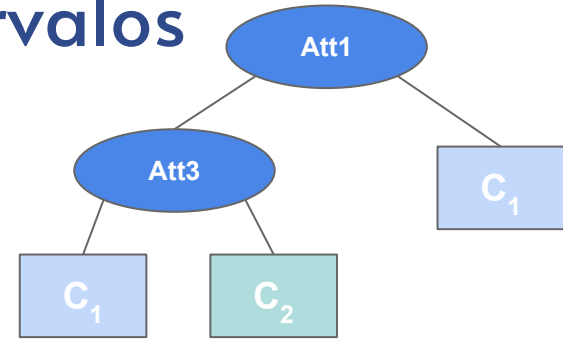
Pausa para a Random Forest

Algoritmos baseados em intervalos

Pausa para a Random Forest

id	Att1	Att2	Att3	Att4	Classe
1					
2					
3					
4					
5					

Out-of-bag

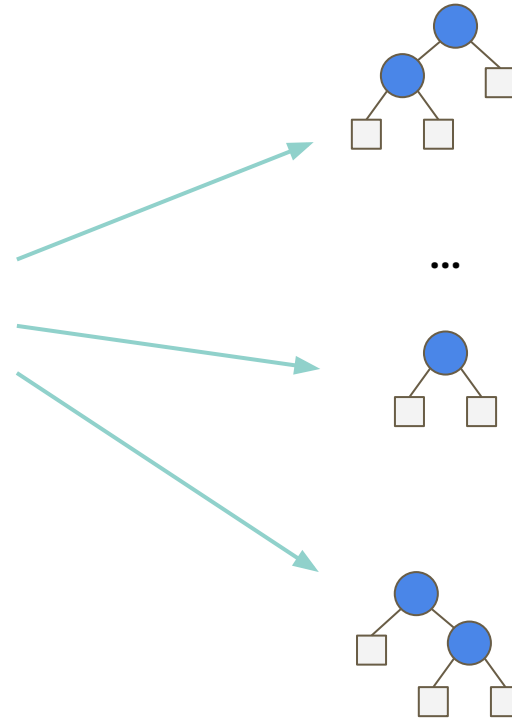


id	Att1	Att3	Classe
1			
4			
3			
3			
1			

Algoritmos baseados em intervalos

Pausa para a Random Forest

id	Att1	Att2	Att3	Att4	Classe
1					
2					
3					
4					
5					



Algoritmos baseados em intervalos

Podemos imaginar um algoritmo que extraia características das séries e utilize qualquer algoritmo “clássico” de AM

Os algoritmos baseados em intervalo fazem isso, porém extraem características de subsequências aleatórias das séries.

Algoritmos baseados em intervalos

Exs:

- Time Series Forest (TSF)
 - Extraem média, desvio padrão e *slope* de cada intervalo
- Random Interval Spectral Ensemble (RISE)
 - Adiciona características de frequência e autocorrelação
- Canonical Interval Forest (CIF)
 - Utiliza as características do *catch22*

Algoritmos baseados em intervalos

O Diverse Representation Canonical Interval Forest (DrCIF) utiliza as mesmas características que o CIF, porém as extrai também na derivativa da série e em seu periodograma (não cabem nesse curso).

Apesar de um pouco lento, obtivemos o melhor resultado “geral” em regressão extrínseca com esse algoritmo.

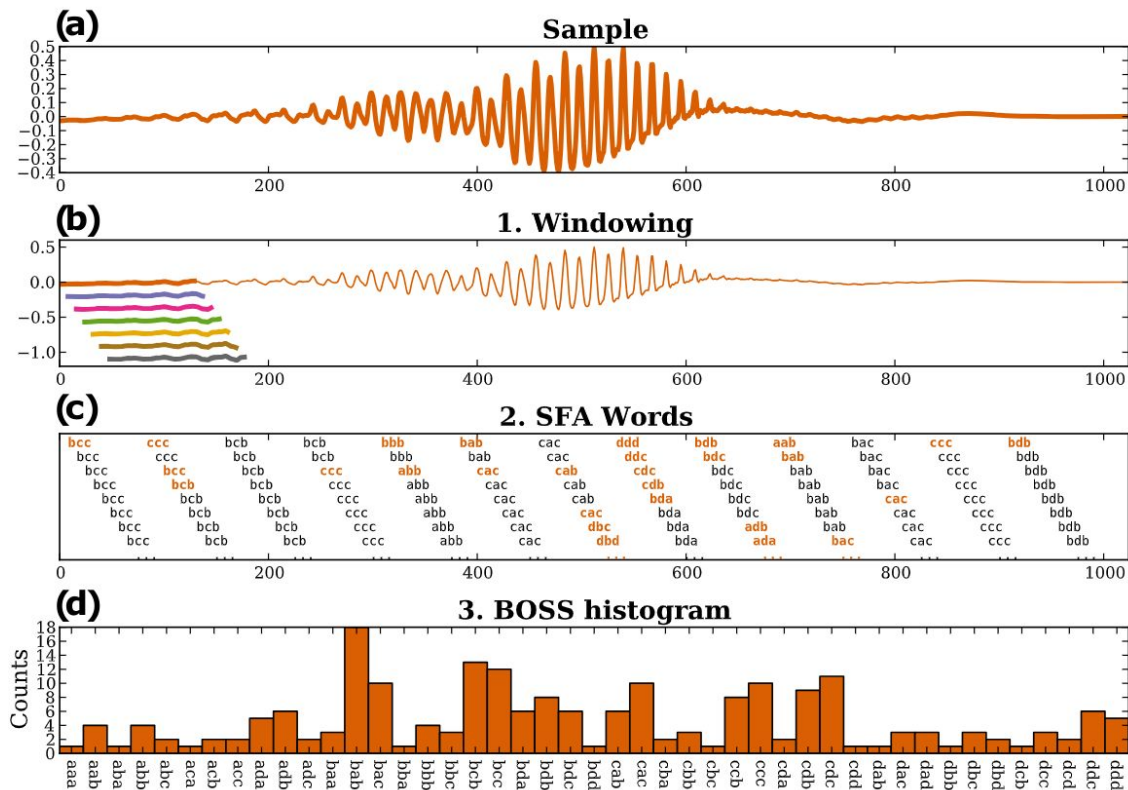
Dicionários

Algoritmos baseados em dicionários

Algoritmos baseados em extração de características inspirada na representação textual *bag-of-words* e similares

Ex: Bag of Symbolic Fourier Approximation Symbols (BOSS)

Algoritmos baseados em dicionários



Algoritmos baseados em dicionários

Nota importante:

- Um algoritmo intervalar pode muito bem usar dicionários

ROCKET
(não está em uma
categoria específica)

RandOm Convolutional KErnel Transform

Está próximo aos baseados em dicionário, mas não é *bag-of-wordish*

- O ROCKET usa kernels convolucionais aleatórios
 - em outras palavras, filtros aleatórios
- Como cada kernel gera um “nova série”, é preciso agregar
 - Max e ppv

E a interpretabilidade ó...



Combinação/comitês

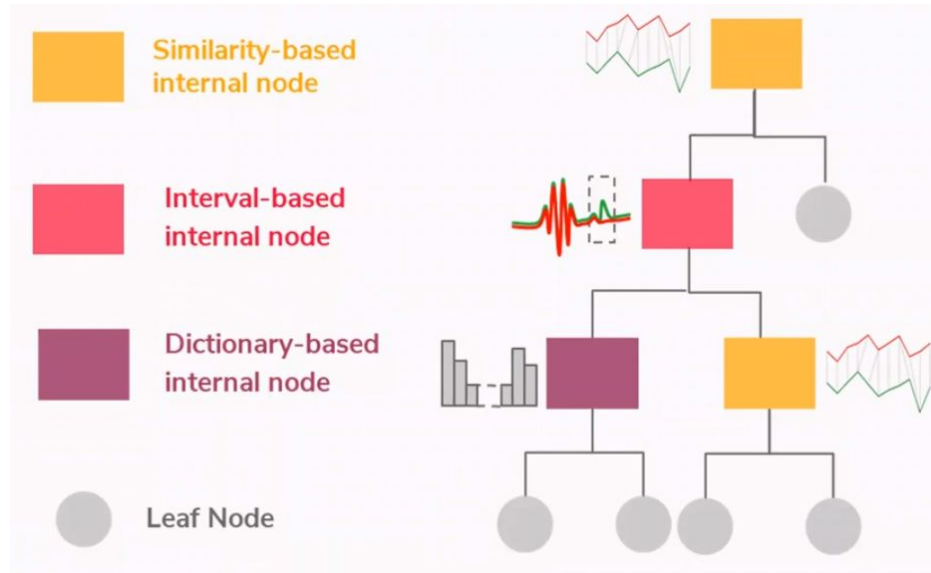
Algoritmos baseados em comitês

“Não há almoço grátis”

Então, vamos combinar :)

Algoritmos baseados em comitês

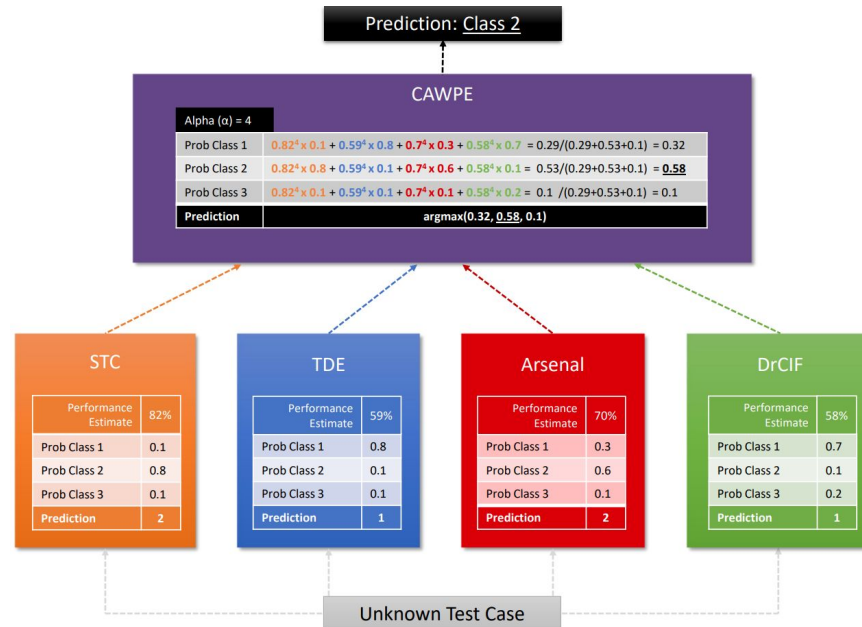
Time Series Combination of Heterogeneous and Integrated Embedding Forest (TS-CHIEF)



<https://youtu.be/SOnHXymw48k>

Algoritmos baseados em comitês

Hierarchical Vote Collective of Transformation-based Ensembles 2
(HIVE-COTE 2.0) - *zoom no próximo slide*



Prediction: Class 2

CAWPE

Alpha (α) = 4	
Prob Class 1	$0.82^4 \times 0.1 + 0.59^4 \times 0.8 + 0.7^4 \times 0.3 + 0.58^4 \times 0.7 = 0.29 / (0.29 + 0.53 + 0.1) = 0.32$
Prob Class 2	$0.82^4 \times 0.8 + 0.59^4 \times 0.1 + 0.7^4 \times 0.6 + 0.58^4 \times 0.1 = 0.53 / (0.29 + 0.53 + 0.1) = \underline{0.58}$
Prob Class 3	$0.82^4 \times 0.1 + 0.59^4 \times 0.1 + 0.7^4 \times 0.1 + 0.58^4 \times 0.2 = 0.1 / (0.29 + 0.53 + 0.1) = 0.1$
Prediction	$\text{argmax}(0.32, \underline{0.58}, 0.1)$

A
Hi
(T

2

STC	
Performance Estimate	82%
Prob Class 1	0.1
Prob Class 2	0.8
Prob Class 3	0.1
Prediction	2

TDE	
Performance Estimate	59%
Prob Class 1	0.8
Prob Class 2	0.1
Prob Class 3	0.1
Prediction	1

Arsenal	
Performance Estimate	70%
Prob Class 1	0.3
Prob Class 2	0.6
Prob Class 3	0.1
Prediction	2

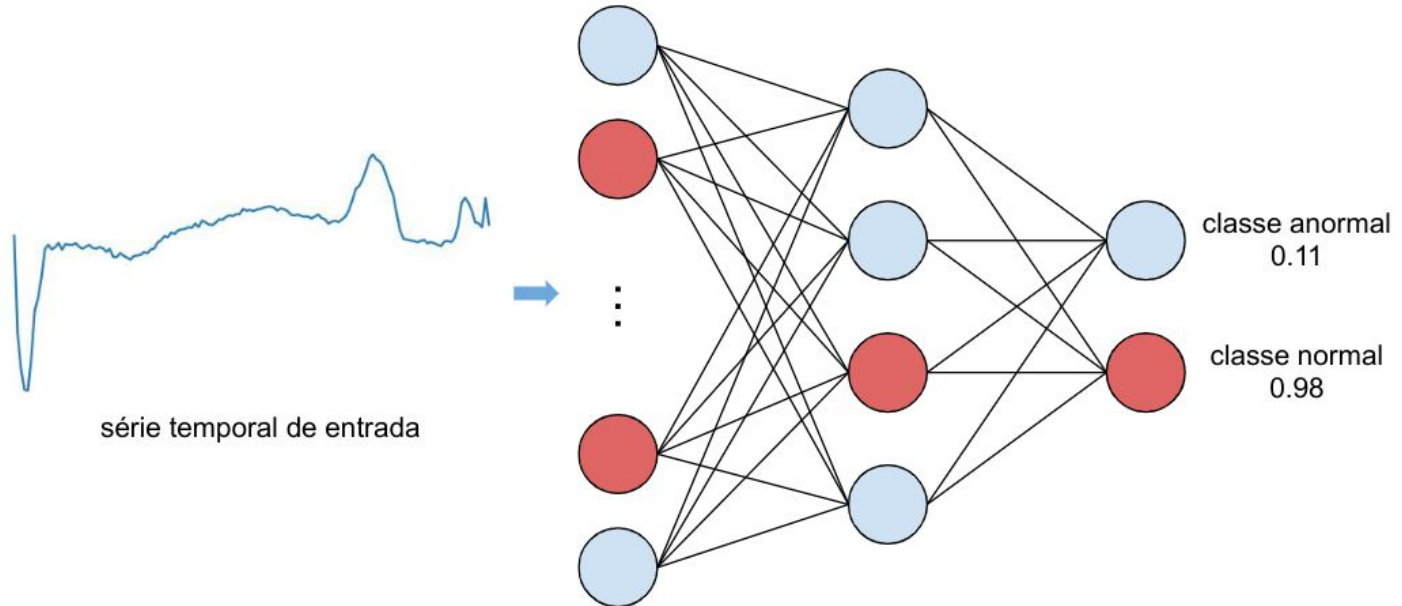
DrCIF	
Performance Estimate	58%
Prob Class 1	0.7
Prob Class 2	0.1
Prob Class 3	0.2
Prediction	1

Unknown Test Case

Aprendizado profundo (*deep learning*)

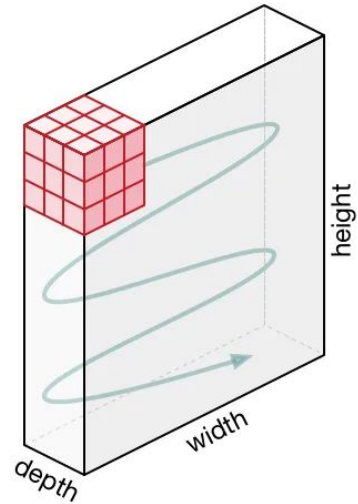
Algoritmos baseados em *deep learning*

Uma extensão das clássicas redes neurais artificiais



Algoritmos baseados em *deep learning*

Convolução



Algoritmos baseados em *deep learning*

Convolução

1	0	1
0	1	0
1	0	1

1 _{x1}	1 _{x0}	1 _{x1}	0	0
0 _{x0}	1 _{x1}	1 _{x0}	1	0
0 _{x1}	0 _{x0}	1 _{x1}	1	1
0	0	1	1	0
0	1	1	0	0

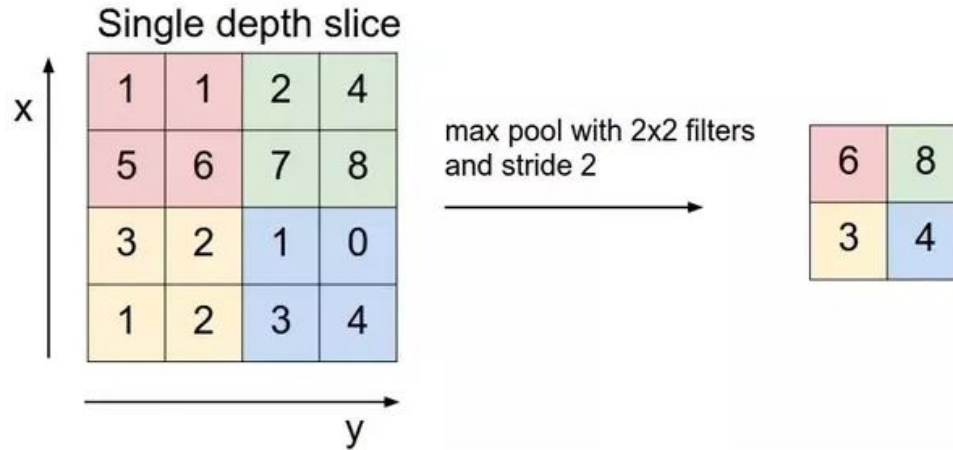
Image

4		

Convolved
Feature

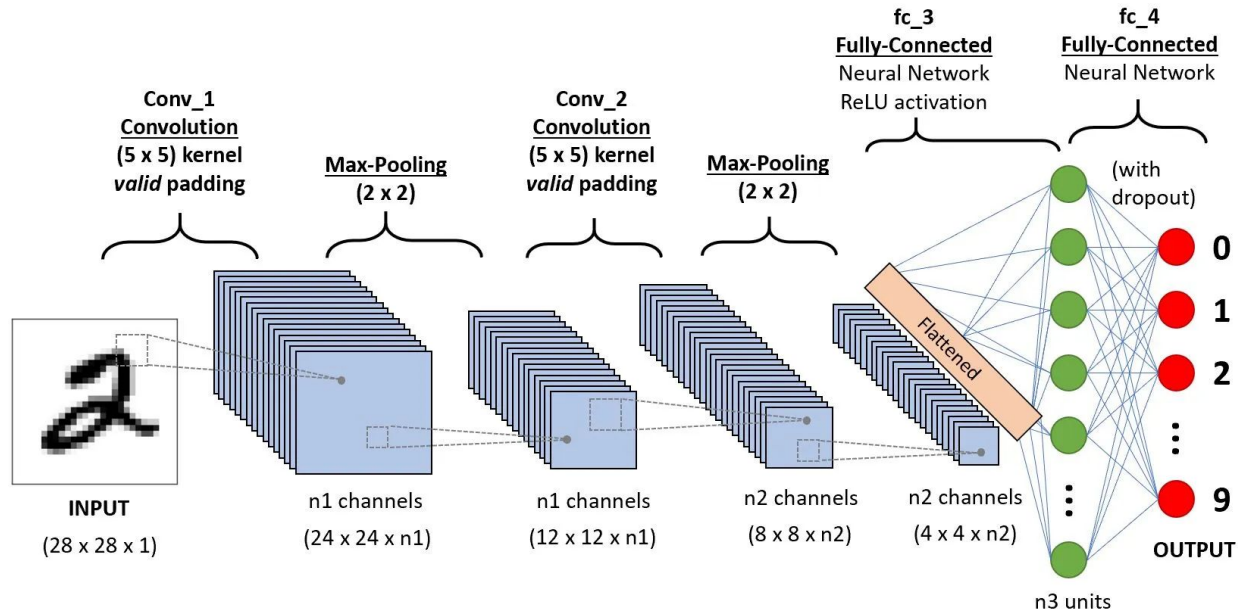
Algoritmos baseados em *deep learning*

Redução de dimensionalidade (pooling)



Algoritmos baseados em *deep learning*

Convolutional Neural Network (CNN)

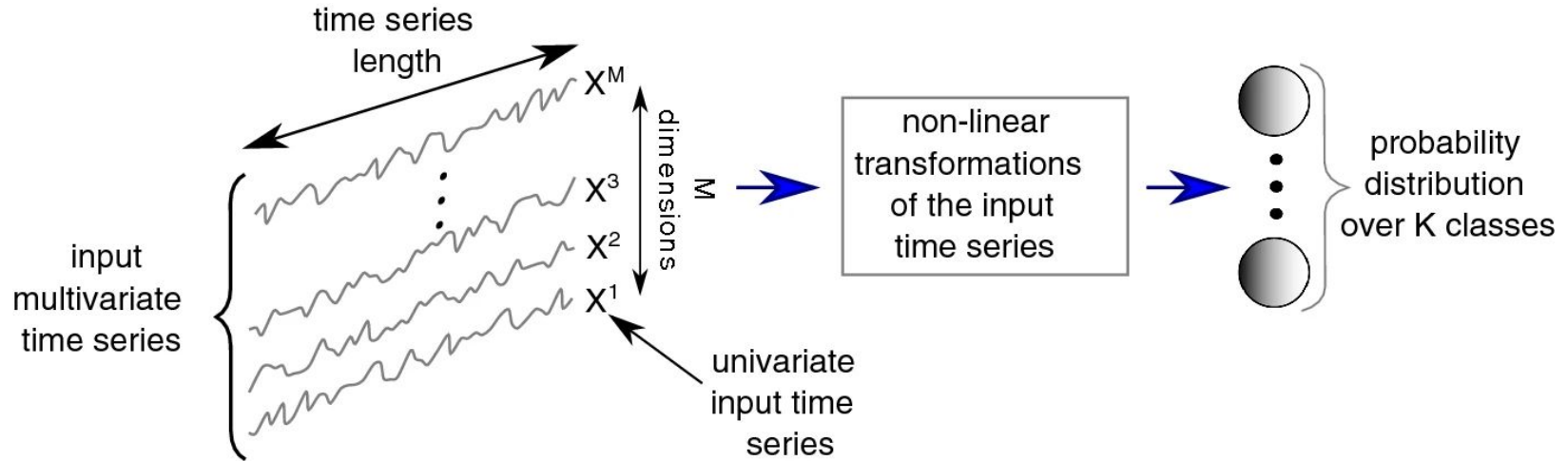


Algoritmos baseados em *deep learning*

Observação importante: RNN vs. CNN

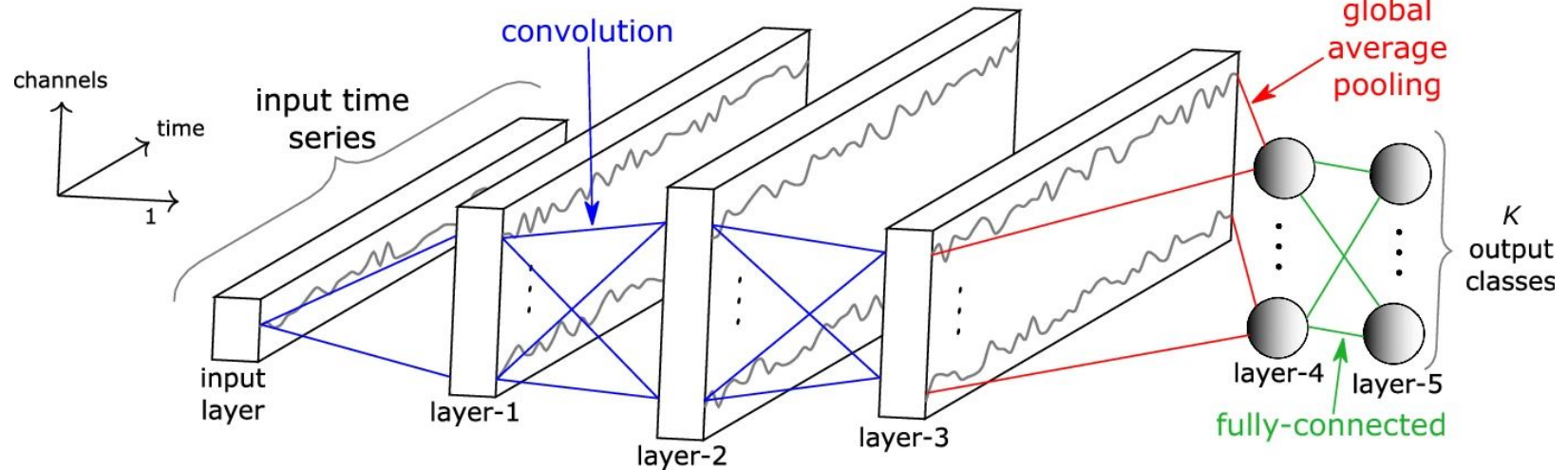
Algoritmos baseados em *deep learning*

Esquema geral da classificação de séries com *deep learning*



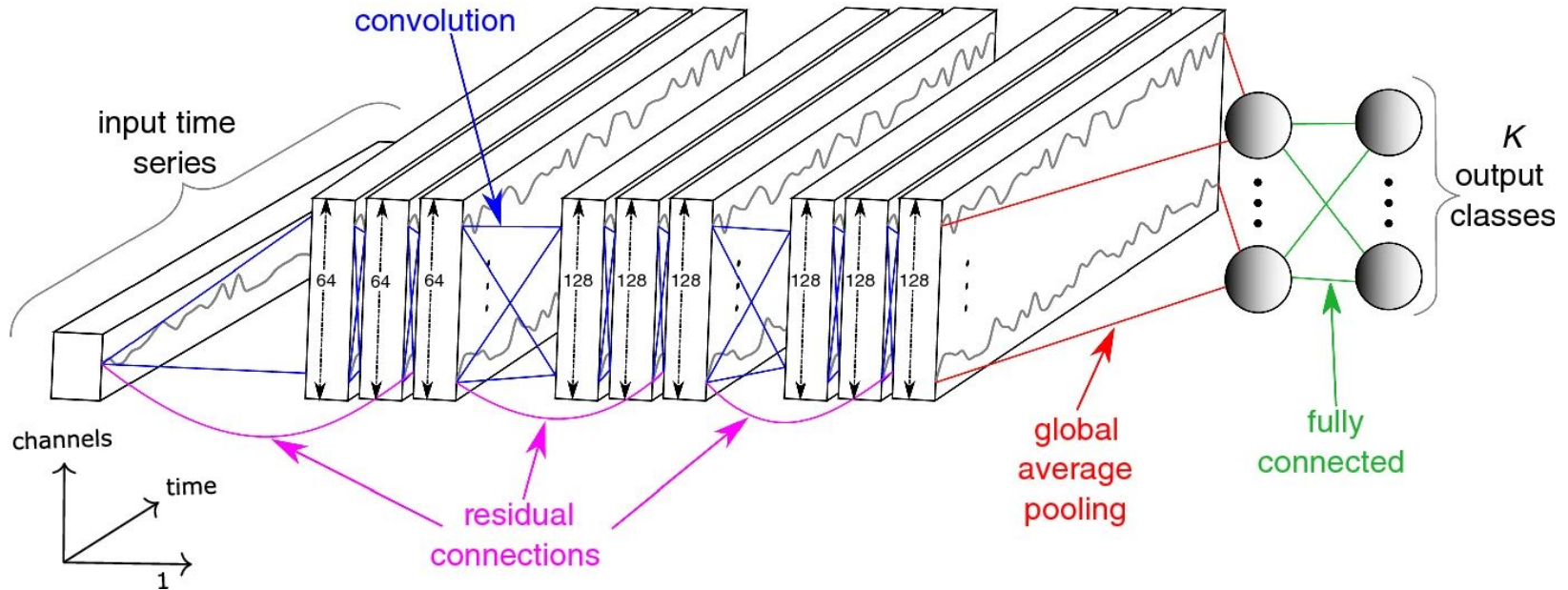
Algoritmos baseados em *deep learning*

Fully-Convolutional Network



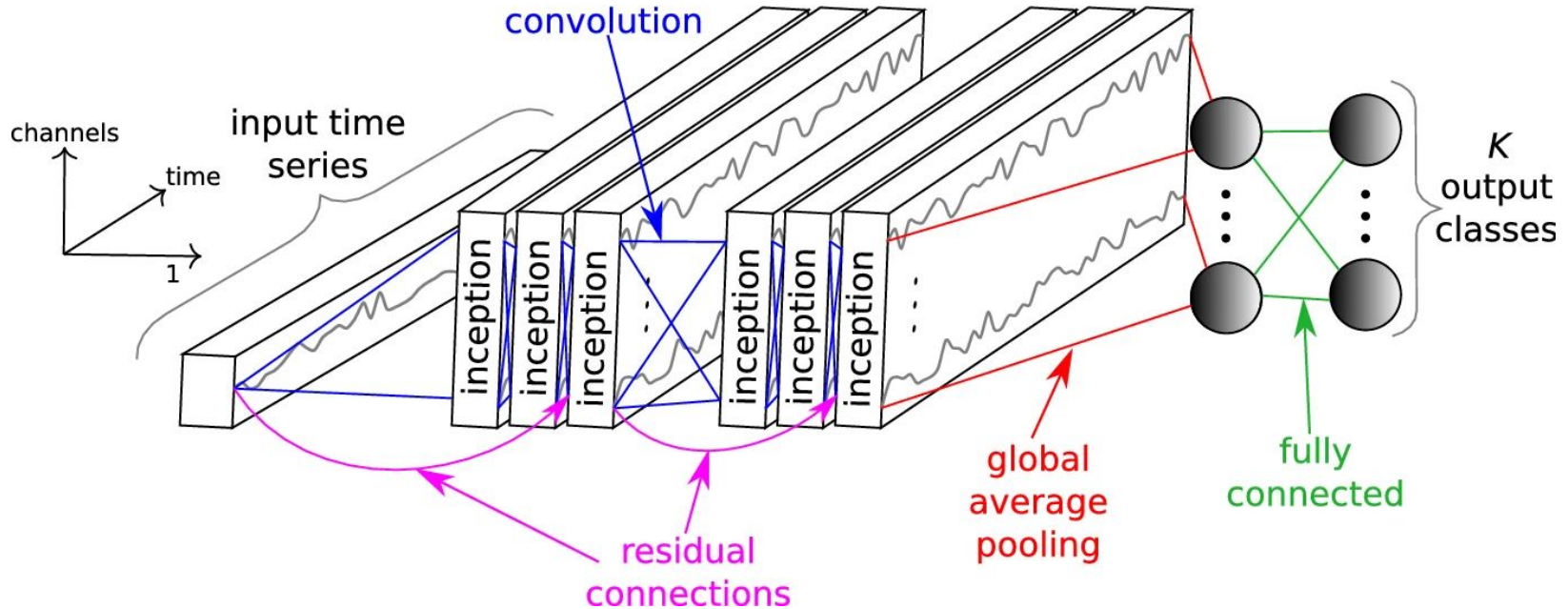
Algoritmos baseados em *deep learning*

ResNet



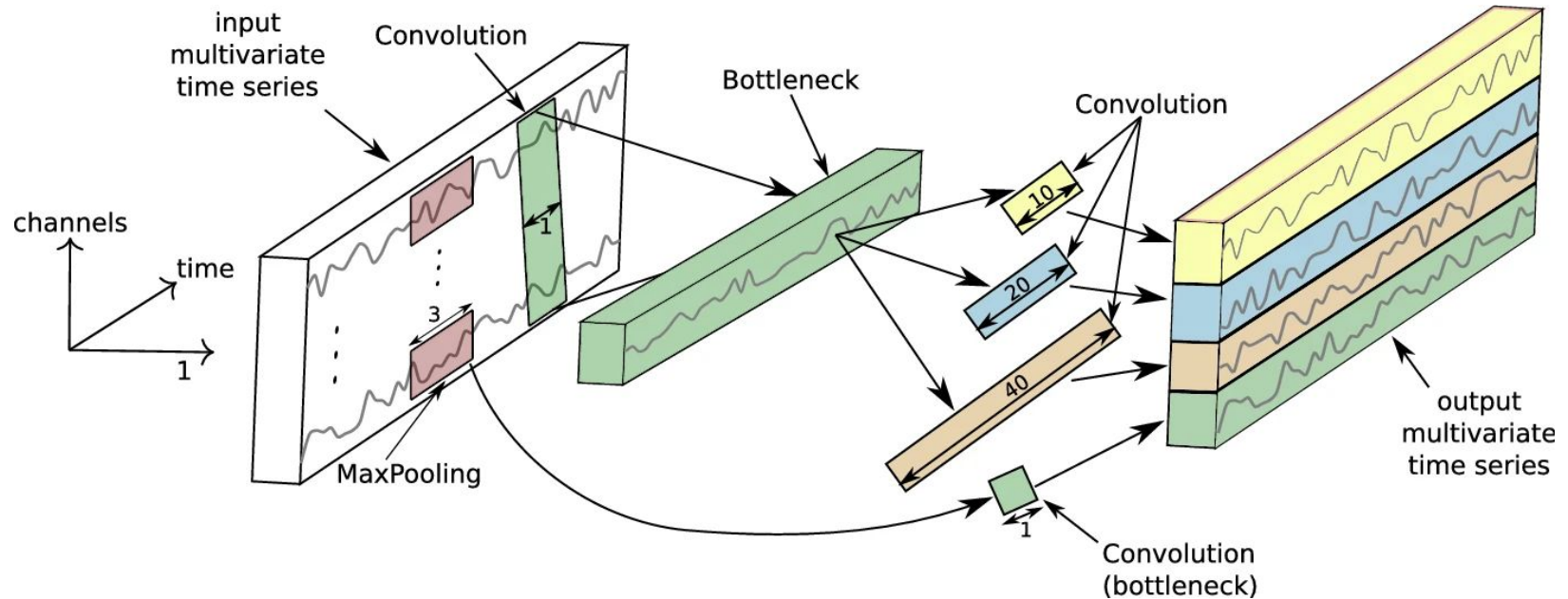
Algoritmos baseados em *deep learning*

InceptionTime



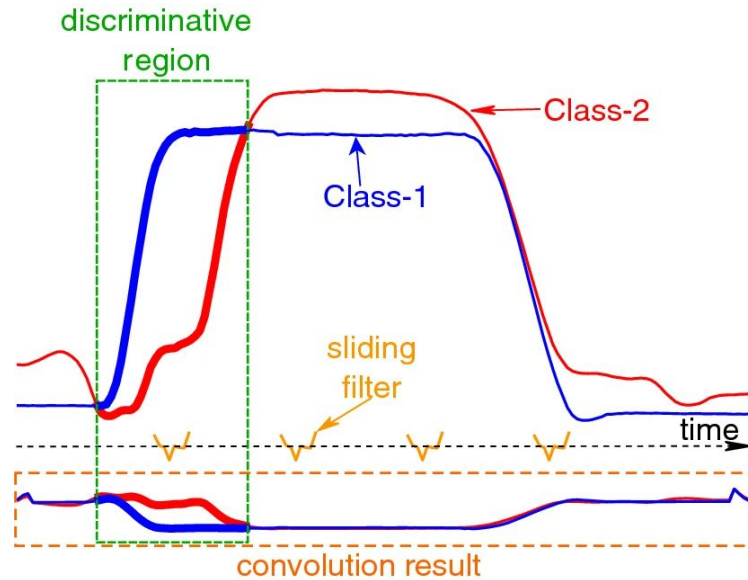
Algoritmos baseados em *deep learning*

InceptionTime



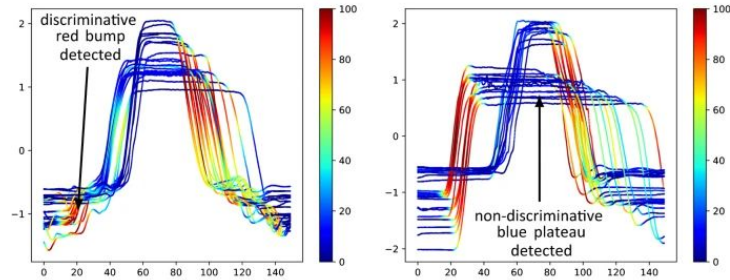
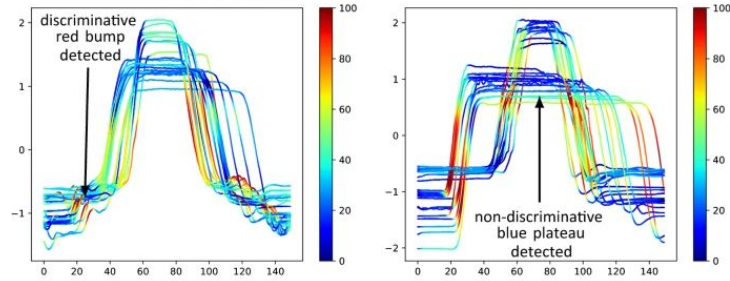
Algoritmos baseados em *deep learning*

Dá para interpretar alguma coisa?



Algoritmos baseados em *deep learning*

Dá para interpretar alguma coisa?



Algoritmos baseados em *deep learning*

Discussão:

- Vantagens e desvantagens
- Arquitetura vs arquitetura

Exemplos práticos

Prática com *deep learning*

Os exemplos serão demonstrados diretamente no notebook

Obrigado

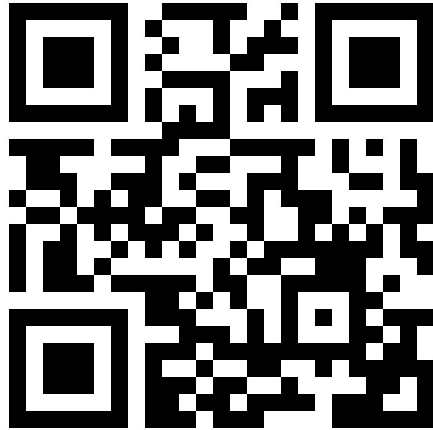
Sinta-se à vontade para me contactar:
diego@usp.br

CREDITS: This presentation template was created by Slidesgo, including icons by Flaticon, and infographics & images by Freepik and illustrations

Este trabalho é apoiado pela

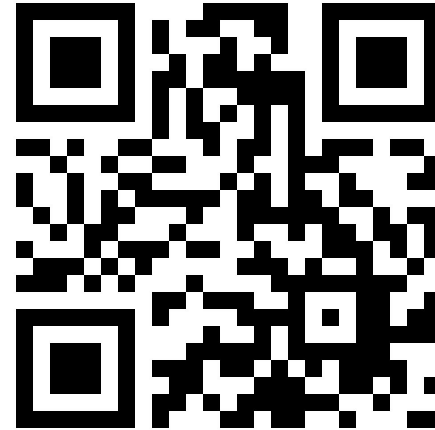


Aprendizado de Máquina Supervisionado para Séries Temporais na Área da Saúde



Link para esta
apresentação

bit.ly/slides-sbcas2023



Link para o Jupyter
notebook com as
demos

bit.ly/colab-sbcas2023