

data

Algoritmos para séries temporais 1

CONTEÚDO

1. Apresentação
2. Definição de Séries Temporais
3. Machine Learning + Pré-processamento de TS
4. Forecasting
5. Algoritmos 1 (Distância, Intervalo, Dicionário)
6. Algoritmos 2 (Deep Learning, Comitês)
7. Aula prática usando Aeon

DIA

26/08 (Merlin)

09/09 (JP)

16/09 (Jorge)

23/09 (Eduarda)

30/09 (Artur)

07/10 (Miller)

14/10 ()



CONTEÚDO

1. Algoritmos de distância
2. Algoritmos de intervalos
3. Algoritmos de dicionários

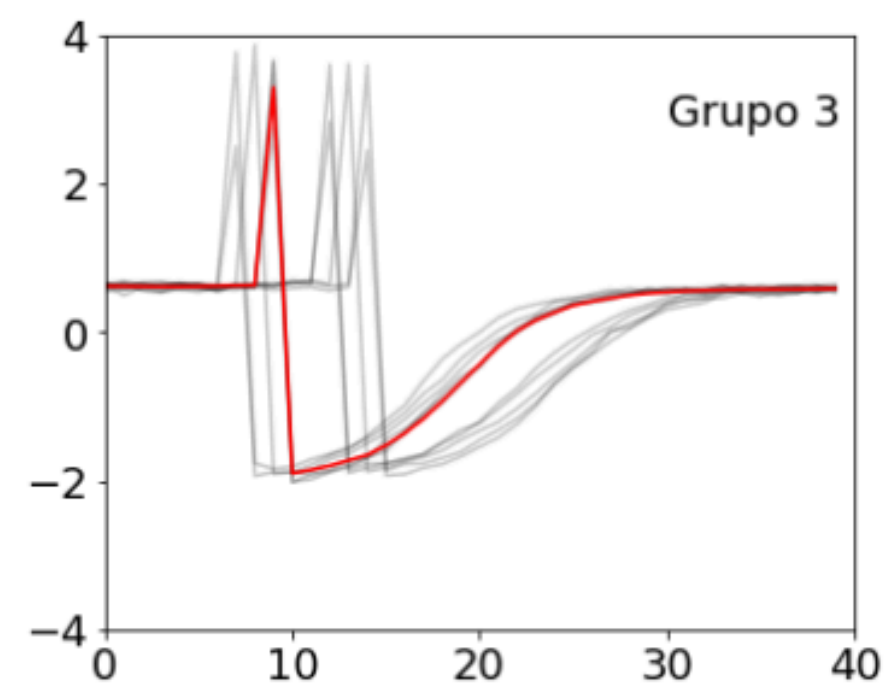
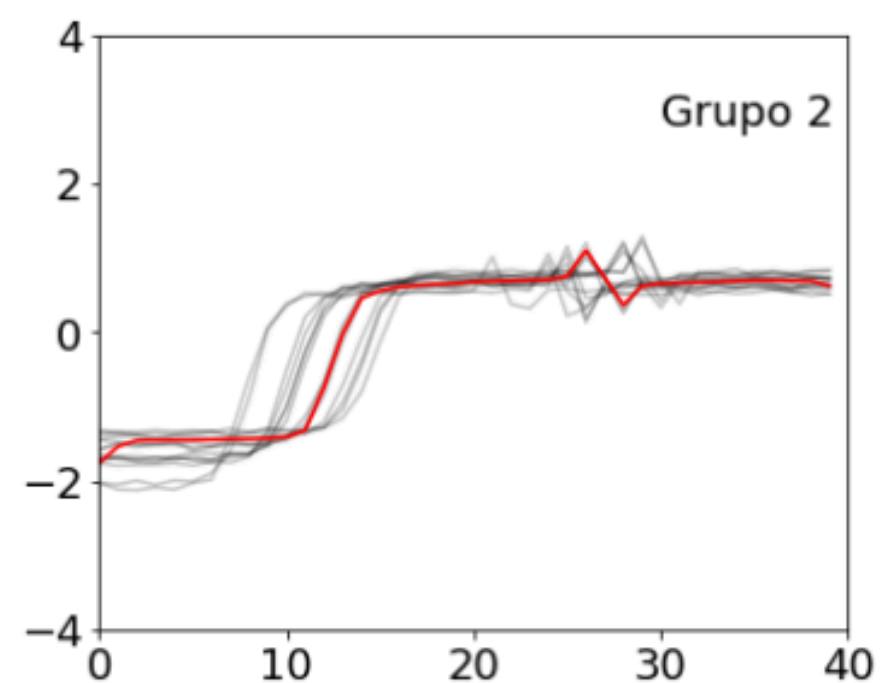
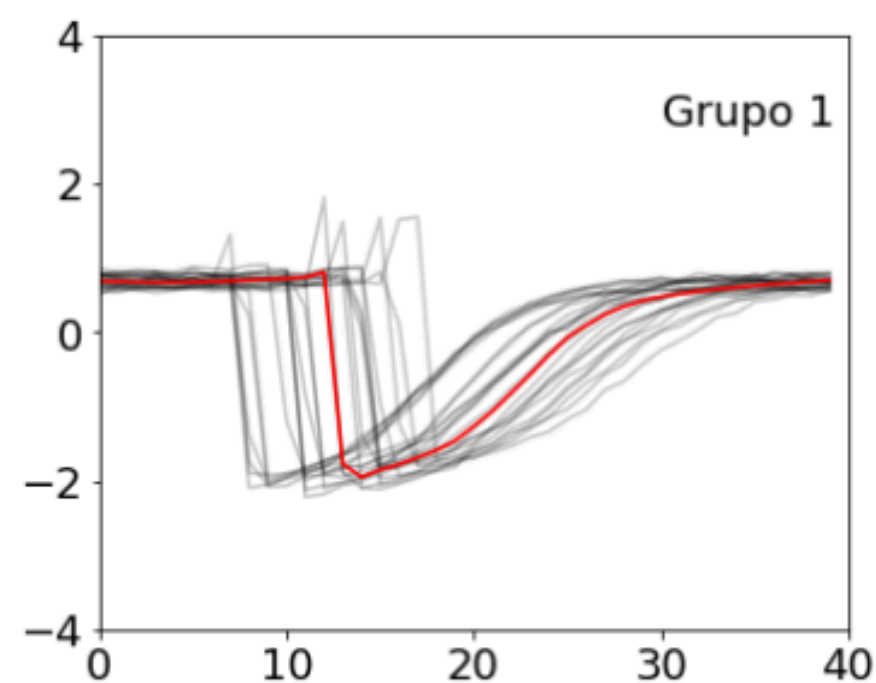


Aqui veremos alguns **algoritmos** mais básicos mas ainda assim clássicos muito utilizados tanto para a classificação quanto a regressão de séries temporais

Ou seja, poderemos compara séries temporais, agrupá-las em classes conhecidas ou não, e até mesmo obter informações de fora da série temporal a partir dela



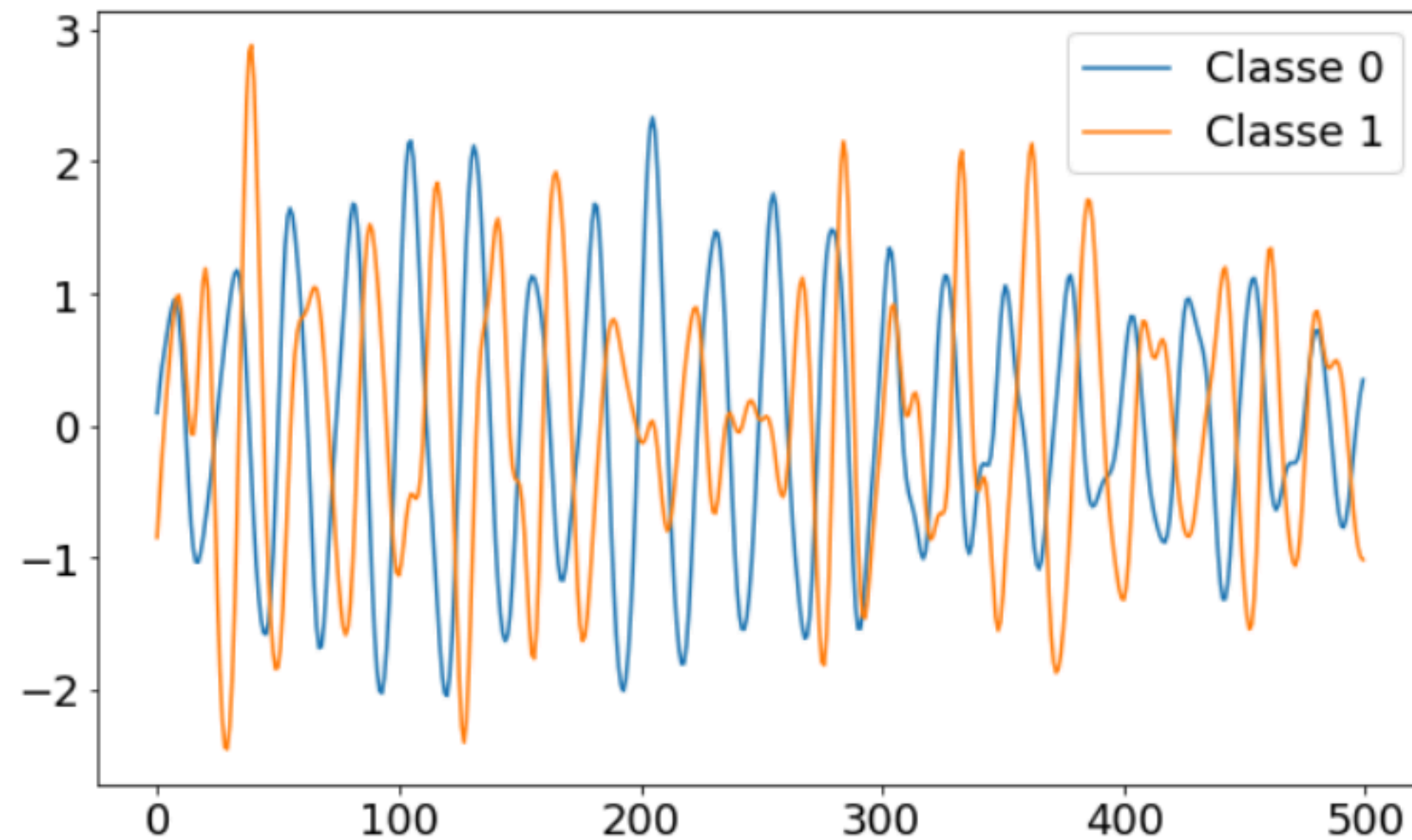
Alguns exemplos



Imagens tiradas do Alura



Alguns exemplos



Imagens tiradas do Alura



Distâncias

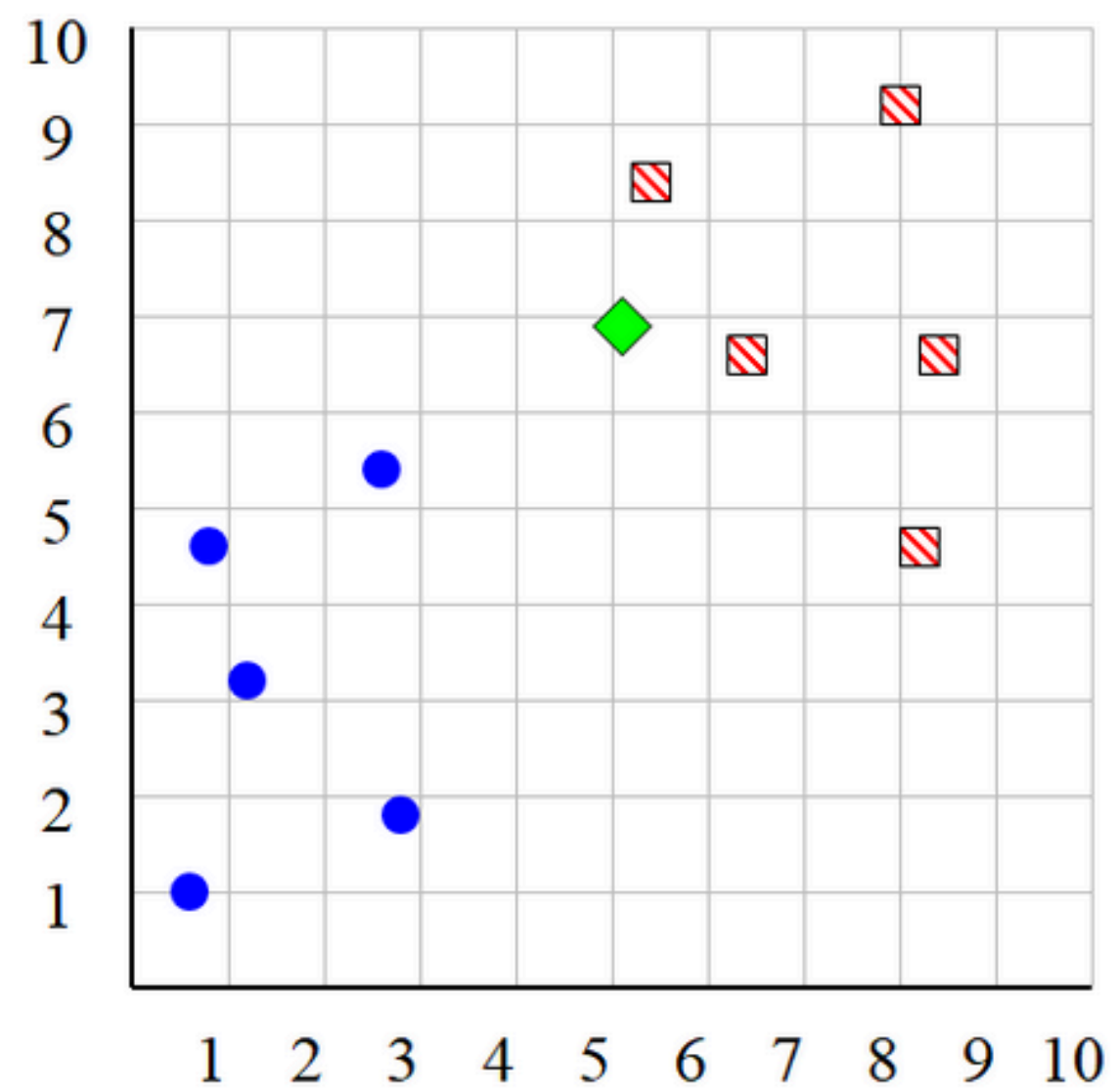
Distância / similaridade

Consiste em avaliar o quanto duas séries temporais de acordo com a distância entre os pontos (menor distância, maior similaridade)

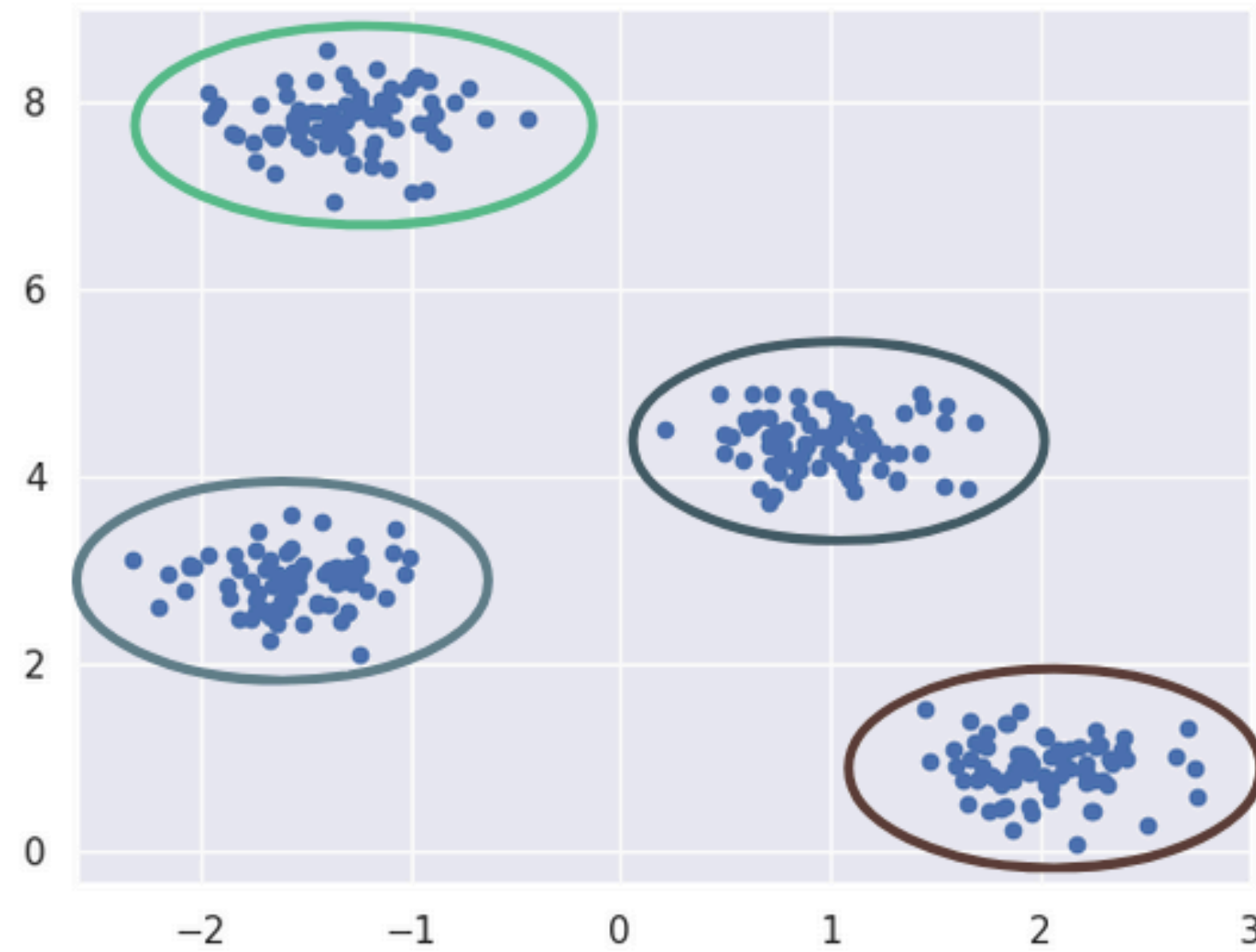
Obtendo uma forma de medir distâncias, podemos utilizar algoritmos clássicos como KNN ou k-means, ou reconhecer padrões nas séries temporais.



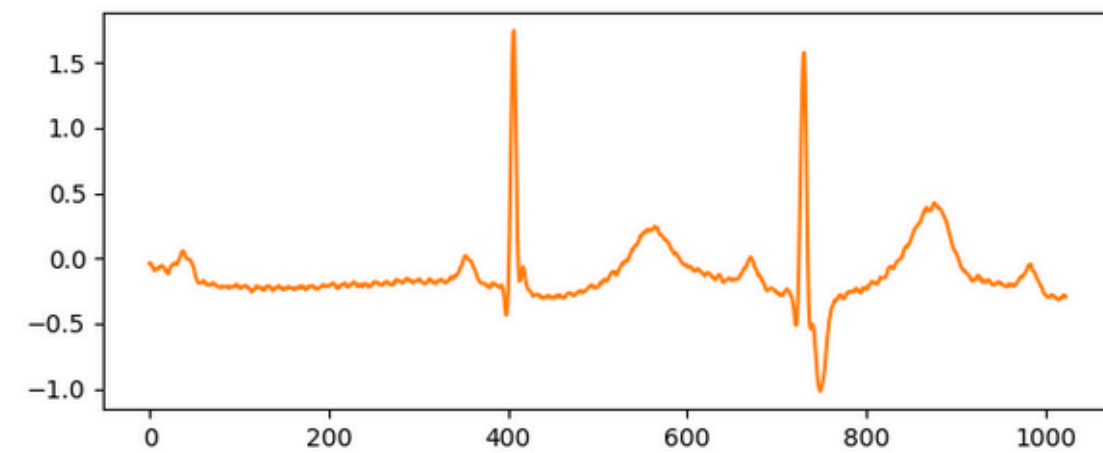
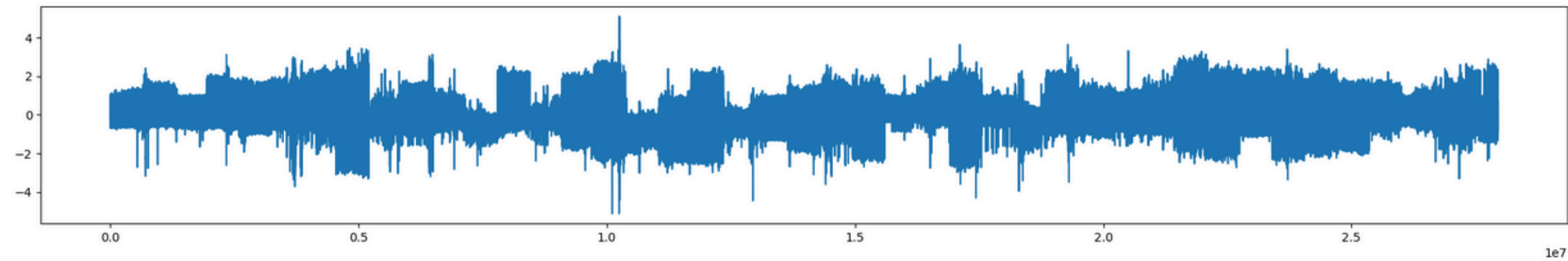
Distância / similaridade



Distância / similaridade



Distância / similaridade

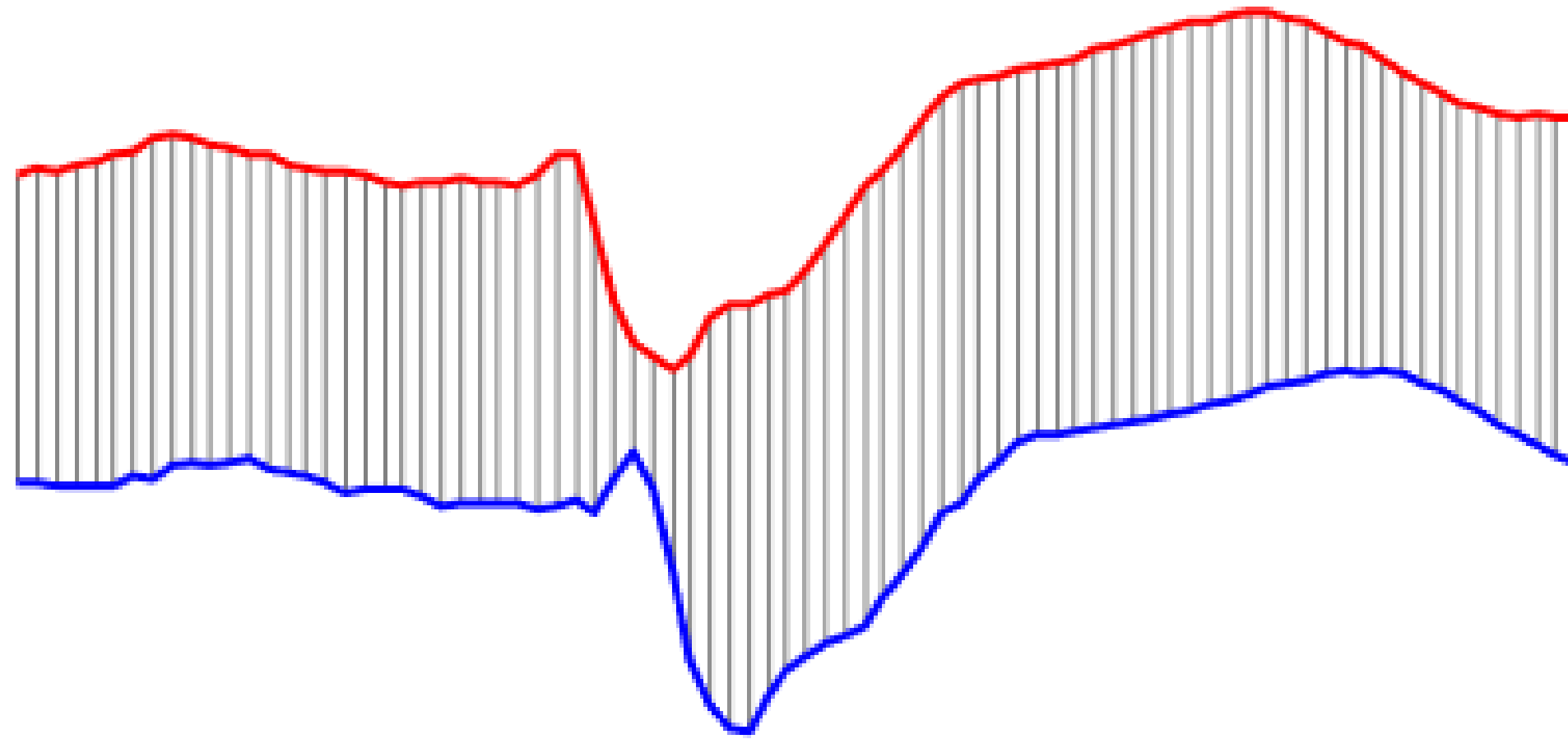


?



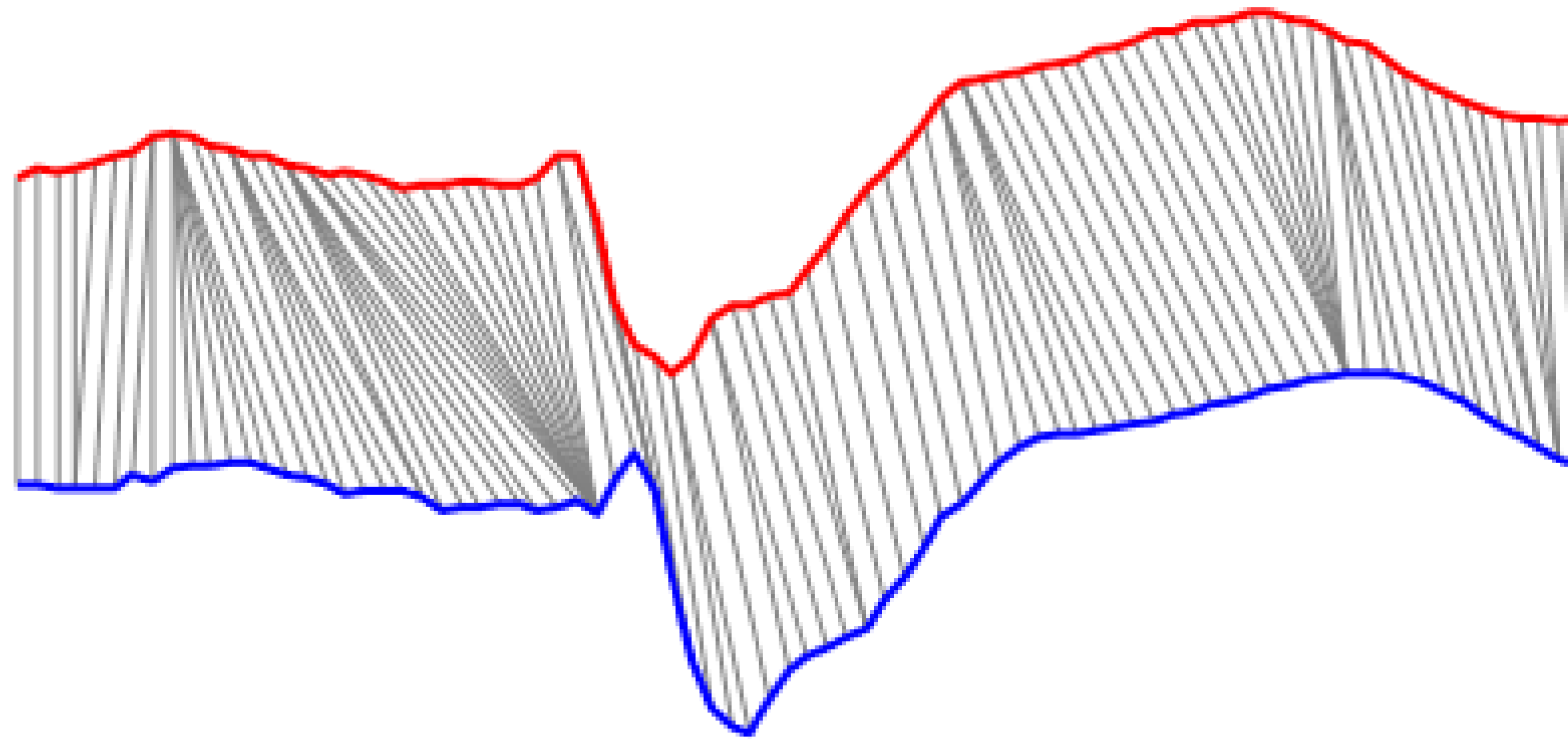
Distância / similaridade

Assim como em aprendizado de máquina clássico, temos várias distâncias diferentes que podemos escolher (manhatan, euclidiana...)



Distância / similaridade


Mas também podemos usar distâncias mais relacionadas às séries temporais, sendo a principal delas a DTW (Distance time warping)



DTW

Essa distância não associa um ponto a um ponto, mas sim um ponto ao ponto com menor distância dentro de um intervalo

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

$(x_i - y_j)^2$ 



DTW

Pairwise Euclidean Distance

0	1	3	2	9	1	14	15	1	2	2	10	7
---	---	---	---	---	---	----	----	---	---	---	----	---

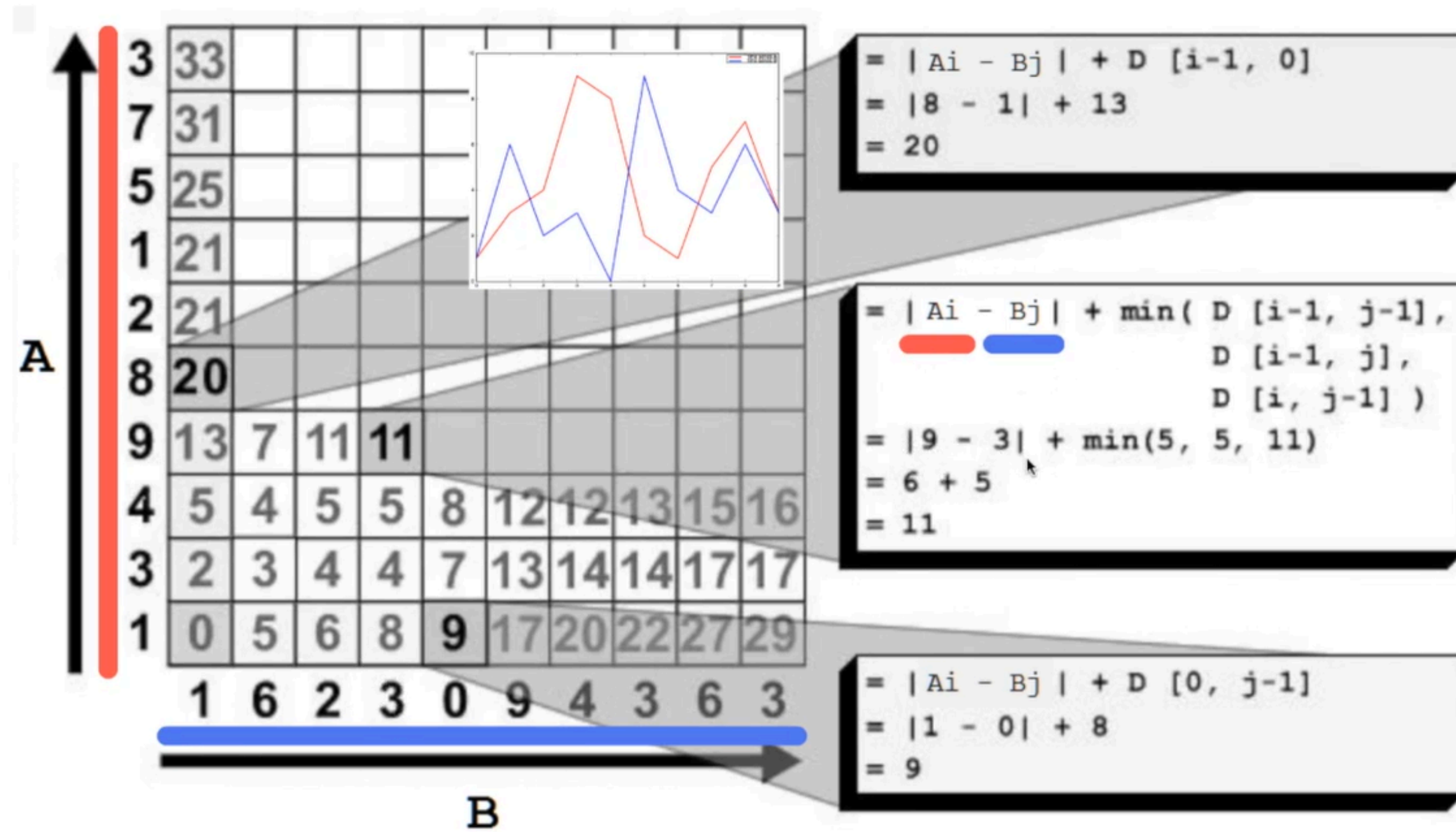
0	1	3	2	9	1	14	15	1	2	2	10	7
---	---	---	---	---	---	----	----	---	---	---	----	---

0.0												
-----	--	--	--	--	--	--	--	--	--	--	--	--

#DistanceProfile

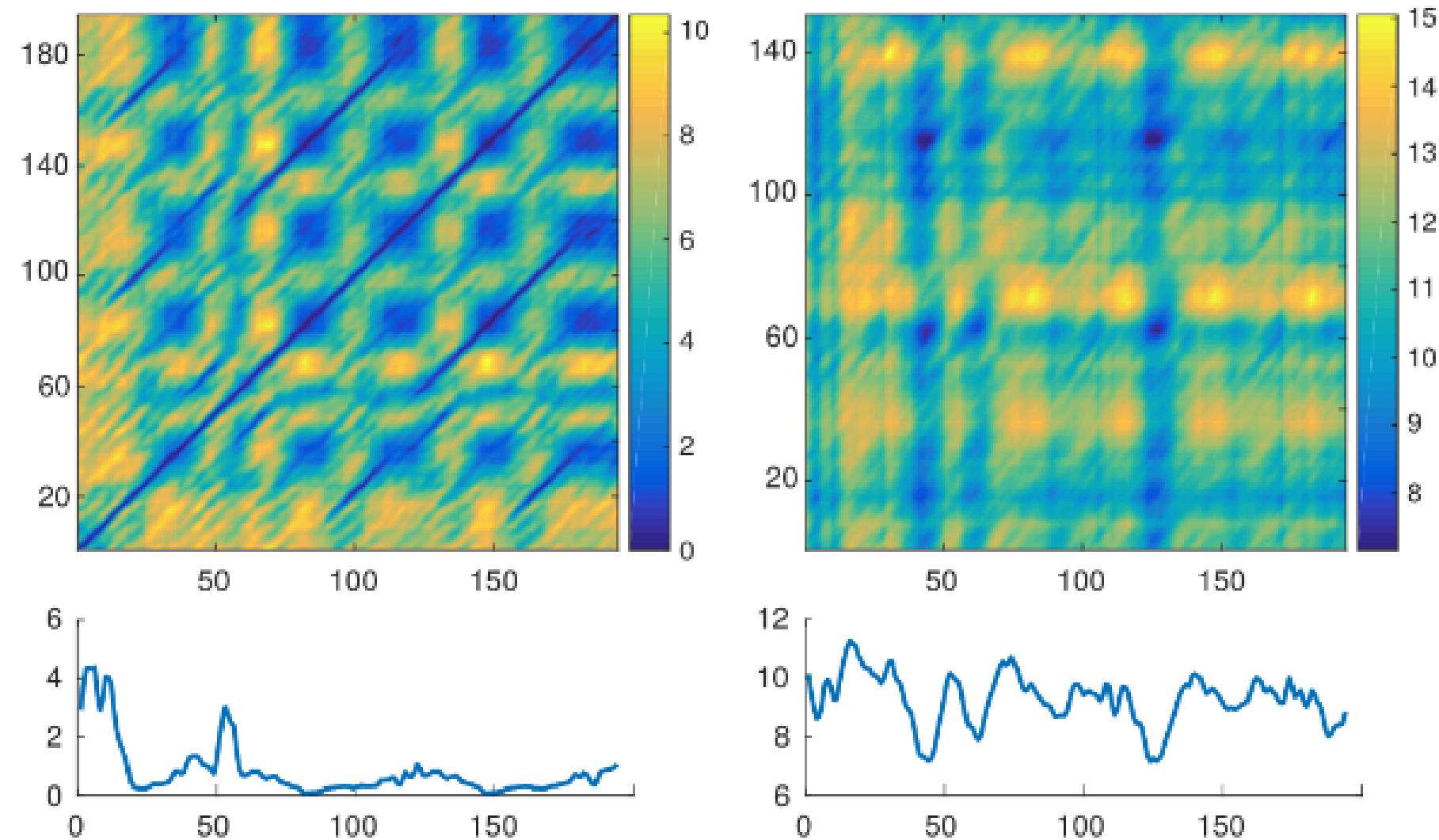


DTW



Bônus: Matrix Profile

Com essas matrizes de distância, entramos em uma outra forma de analisar as séries temporais, as Matrix profile



Bônus: Matrix Profile

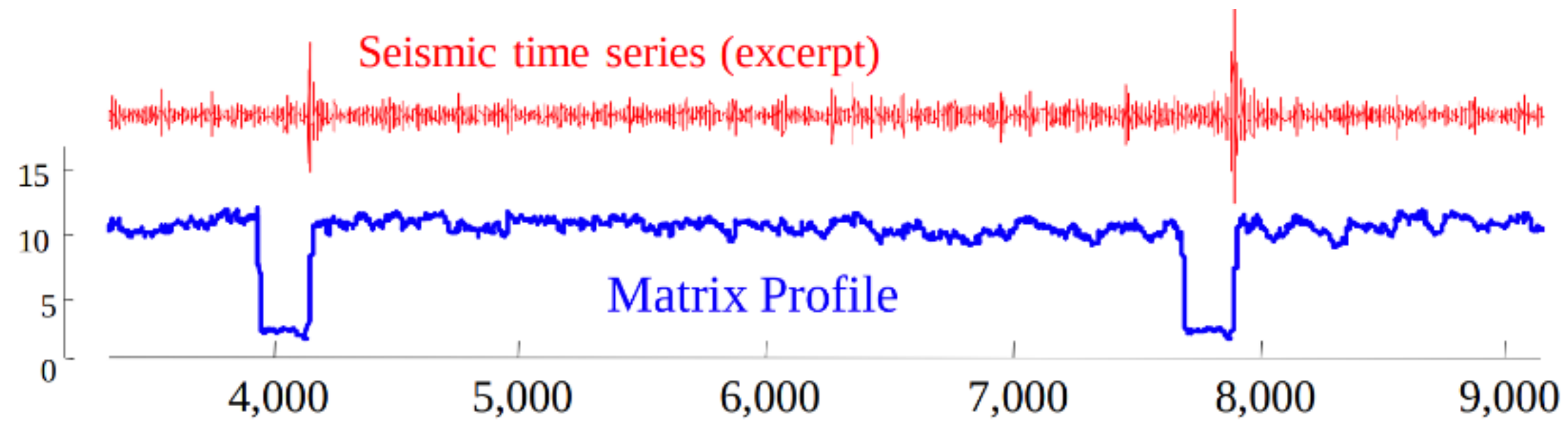
Matrix profile, nada mais é do que a matriz de distância que criamos, podendo ser entre séries diferentes ou comparando uma série consigo mesma

Com o matrix profile, podemos encontrar informações interessantes como motifs(subsequência que se repete) e discords (anomalias)



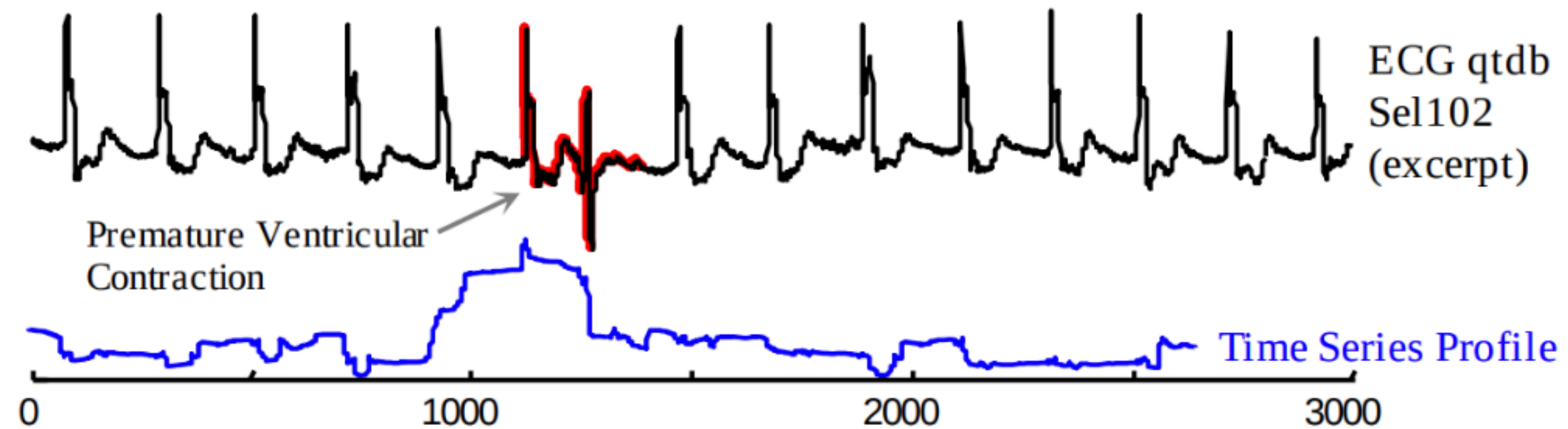
Bônus: Matrix Profile

Exemplo de Motif



Bônus: Matrix Profile

Exemplo de discord





Intervalos



Algoritmos de intervalo

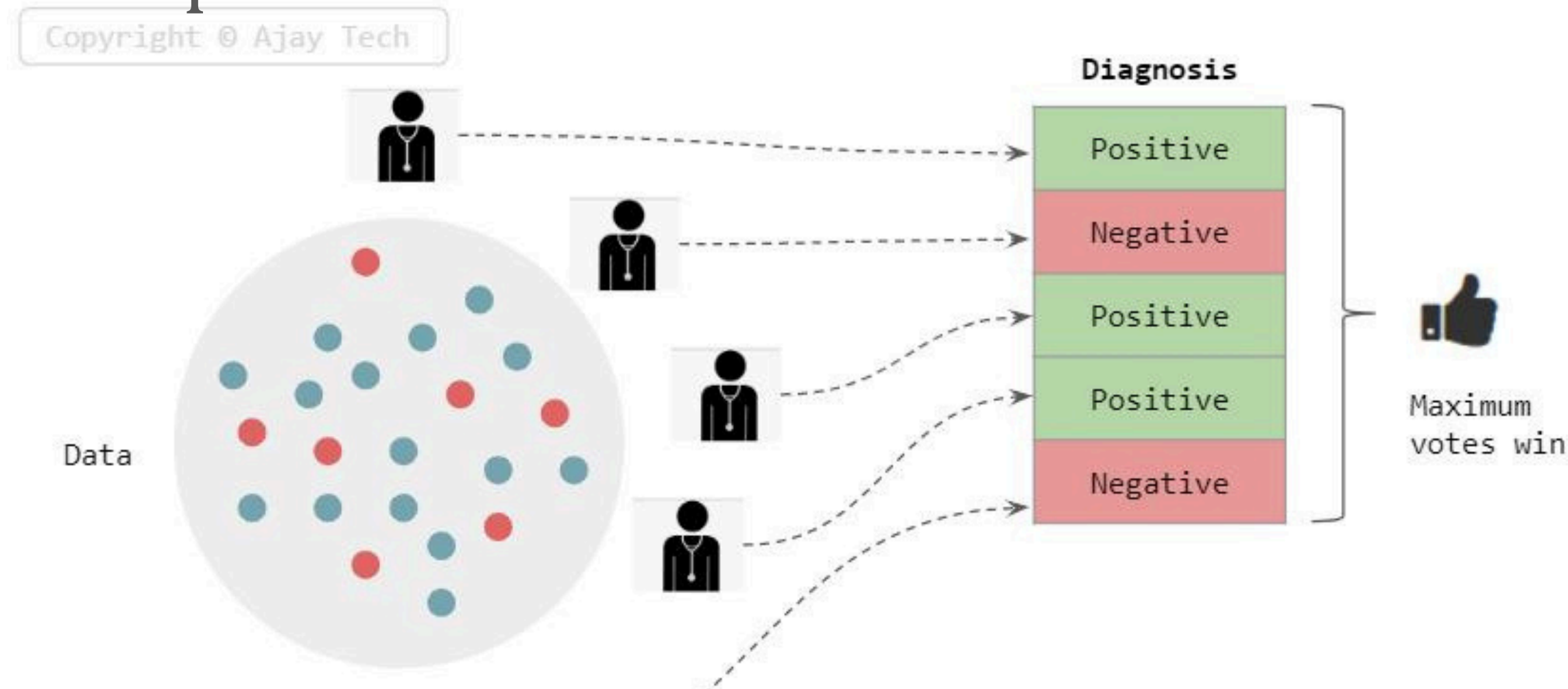
São algoritmos que trabalham com partes da série temporal.

Assim acaba-se treinando vários modelos, um para cada parte e no fim, faz-se um ensemble desses modelos.



Algoritmos de intervalo

Para quem nunca viu, um ensemble é uma forma de juntar resultados de modelos, podendo ser feito usando a média entre eles, votações e por aí vai.



Algoritmos de intervalo

Entre os algoritmos de destaque, temos:

- TSF
- RISE
- CIF



TSF



O TSF ou Time Series Forest é um modelo que aleatoriamente escolhe um intervalo da série e extrae a média, desvio padrão e slope.

Com esses dados, ele constrói uma árvore de decisão



RISE

O RISE ou Random Interval Spectral Ensemble é bem parecido, só que extraindo informações relacionadas à frequência, como o power spectrum e a correlação.



CIF



Assim como os anteriores, o CIF (ou Canonical Interval Forest) constrói uma árvore, só que ele extrai características utilizando o CATCH 22



CAnonical Time-series CHaracteristics ou CATCH 22



Conjunto de 22 características extraídas de uma série temporal após um estudo extensivo em mais de 1000 características.

É consideravelmente rápido, e muito útil para aplicar algoritmos de machine learning em séries temporais



CAnonical Time-series CHaracteristics ou CATCH 22



Algumas features

- DN_HistogramMode_5 - Moda de um histograma com 5 bins.
- DN_HistogramMode_10 - Moda de um histograma com 10 bins.
- SB_BinaryStats_mean_longstretch1 - Comprimento médio das sequências contínuas de 1s em uma série binarizada.
- DN_OutlierInclude_p_001_mdrmd - Proporção de outliers com um limite de 0,001.



CAnonical Time-series CHaracteristics ou CATCH 22



Para saber mais, tem o site deles explicando mais certinho

<https://time-series-features.gitbook.io/catch22-features>



Adendo

Essa ideia de ensemble também pode ser usada em outras tarefas.

Podemos por exemplo juntar vários algoritmos de distância usando distâncias diferentes.

Assim temos algoritmos como o COTE e o HIVE



Dicionários

Dicionários

Por ser uma área relativamente recente, muita coisa das séries temporais vem emprestada de outras áreas da IA.

Seja extrair features e usar técnicas de Machine Learning clássico, usar medidas de distância vindas da análise de áudio com a DTW, ou usar técnicas da visão computacional como veremos mais pra frente em redes neurais.

Os pesquisadores já tentaram de tudo.



Dicionários

Então por que não transformar a série temporal em um texto e aplicar técnica de Processamento de linguagem natural.

Para suprir essa necessidade, temos as técnicas de dicionário, sendo a mais famosa delas a Bag-of-SFA-Symbols (BOSS)

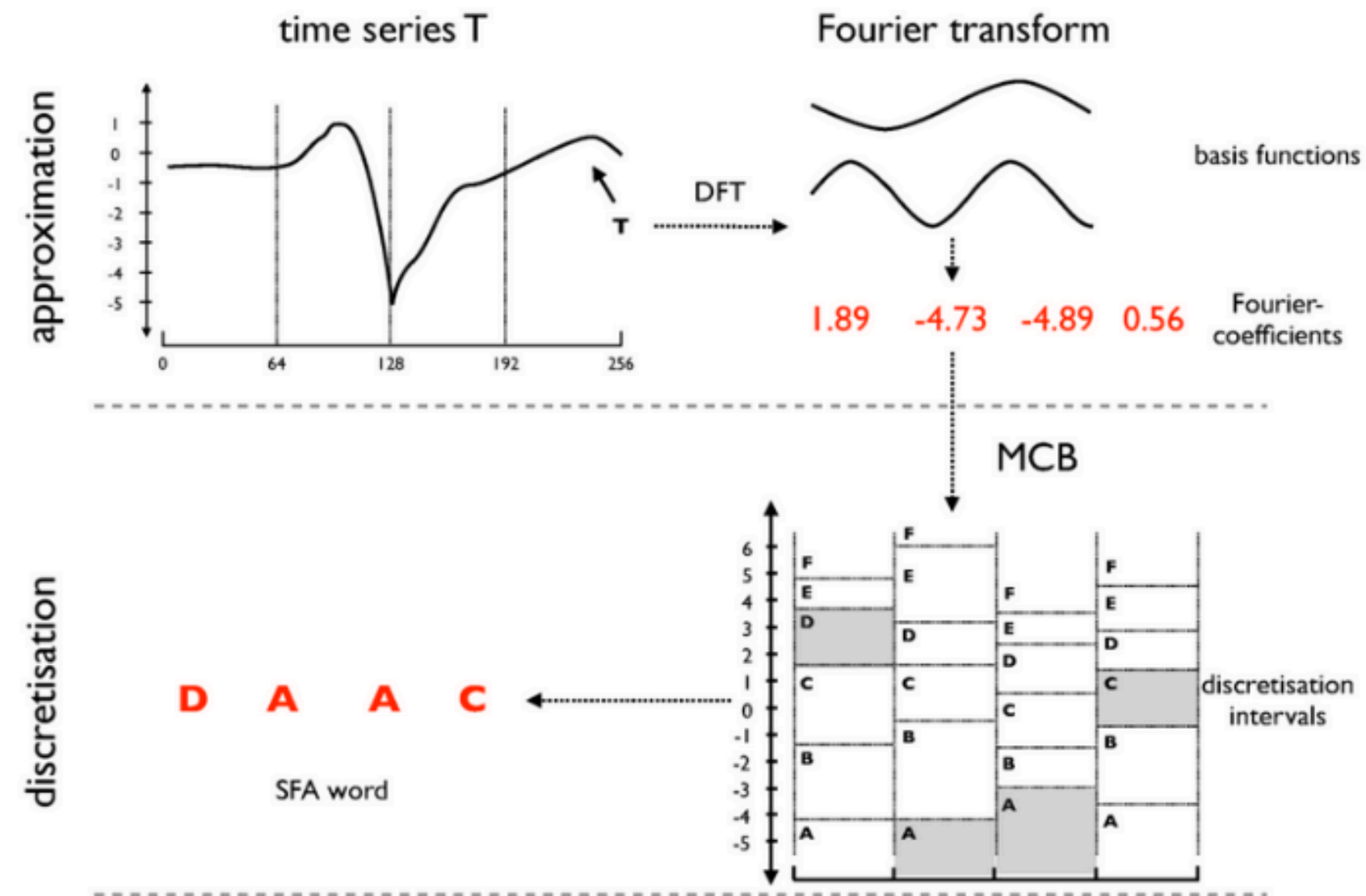


BOSS

O BOSS consiste em segmentar a série temporal e aplicar transformações nesses segmentos usando a Symbolic Fourier Approximation (SFA).

O SFA discretiza os coeficiente obtidos pela transformada de fourier, sendo cada um desses termos discretos uma letra.

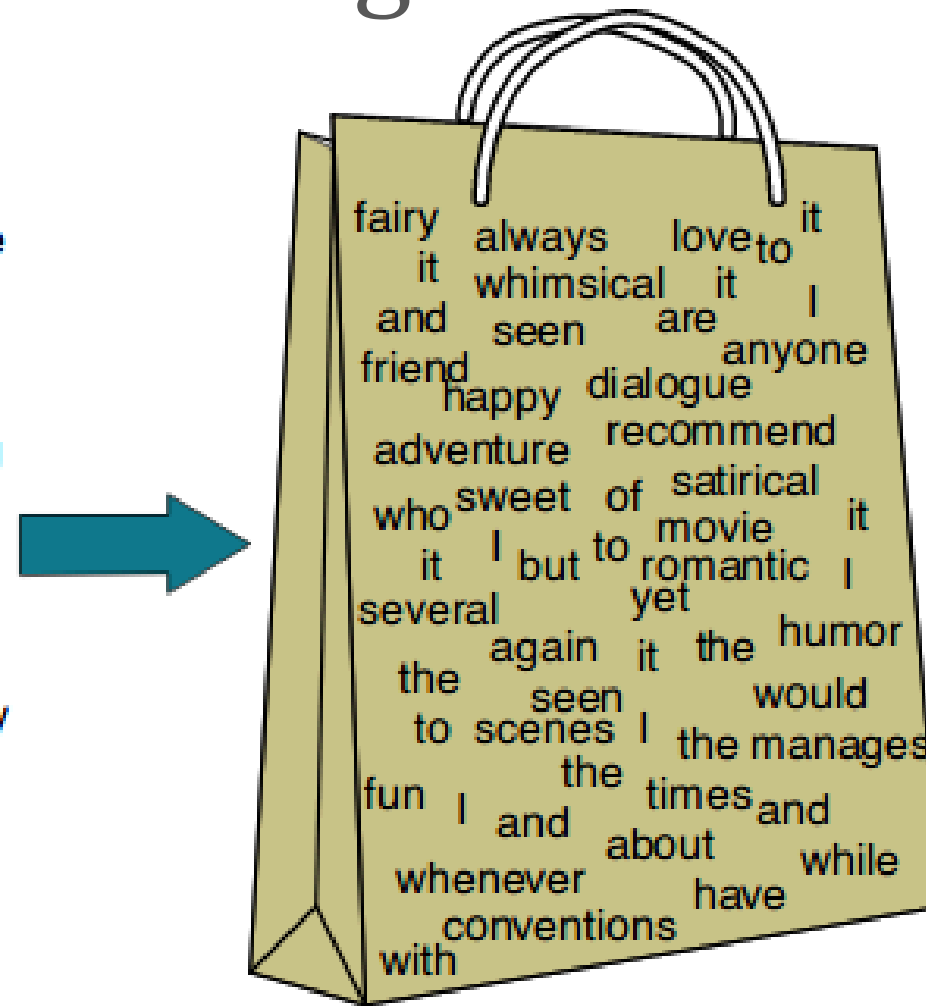
BOSS



BOSS

Assim conseguimos que a série se transforme em sequências de caracteres, e podemos aplicar algoritmos como bag of words nelas

I love this movie! It's sweet,
but with satirical humor. The
dialogue is great and the
adventure scenes are fun...
It manages to be whimsical
and romantic while laughing
at the conventions of the
fairy tale genre. I would
recommend it to just about
anyone. I've seen it several
times, and I'm always happy
to see it again whenever I
have a friend who hasn't
seen it yet!



it	6
I	5
the	4
to	3
and	3
seen	2
yet	1
would	1
whimsical	1
times	1
sweet	1
satirical	1
adventure	1
genre	1
fairy	1
humor	1
have	1
great	1
...	...

Obrigado!

