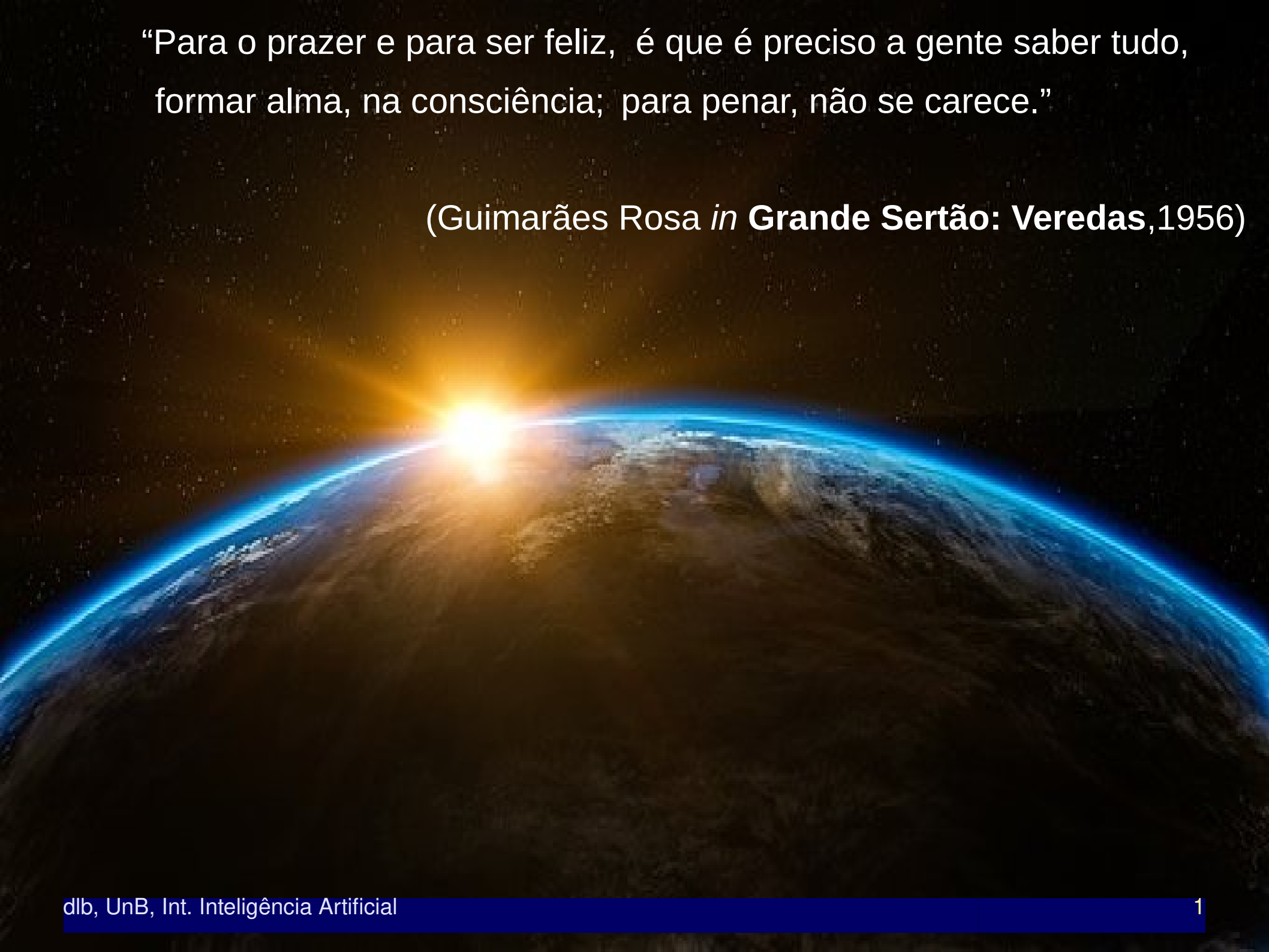


“Para o prazer e para ser feliz, é que é preciso a gente saber tudo,
formar alma, na consciência; para penar, não se carece.”

(Guimarães Rosa *in* **Grande Sertão: Veredas**, 1956)



Introd. Inteligência Artificial

Roteiro da aula:

- ♦ **Abordagem K-nn (Vizinhos mais próximos);**
- ♦ **Exemplos;**
- ♦ **Métricas de similaridade e distâncias em reconhecimento/aprend. de padrões;**

Com slides adaptados de E.Eaton(UPenn)

Abordagem k-nn (k nearest neighbors)

Aprendizagem por instâncias

- Dados não são transformados
- Dados são simplesmente armazenados, e quando um novo exemplo é encontrado, um conjunto de exemplos similares, relacionados, são recuperados da memória e usados para classificar o novo dado.
- Não formam função alvo, simplesmente calculam a classificação de nova instância.

Abordagem k-nn (k nearest neighbors)

Aprendizagem por instâncias

- Aprendizagem Supervisionada
- Não paramétrica

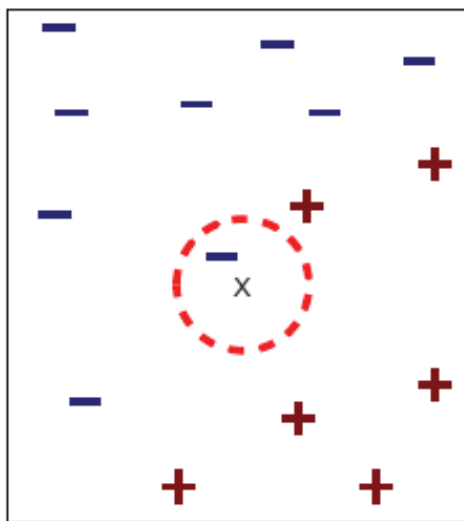
Abordagem k-nn (k nearest neighbors)

Aprendizagem por instâncias

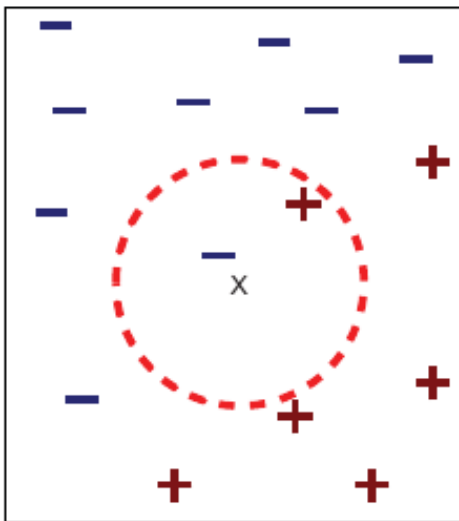
- Modelo mais simples usado como aprendizagem de instâncias.
- K-nn assume que todos os exemplos são pontos em algum espaço n-dimensional e define vizinhos em termos de distância (usualmente Euclideana em \mathbb{R})
- K é o número de vizinhos considerados.

Abordagem k-nn (k nearest neighbors) Aprendizagem por instâncias

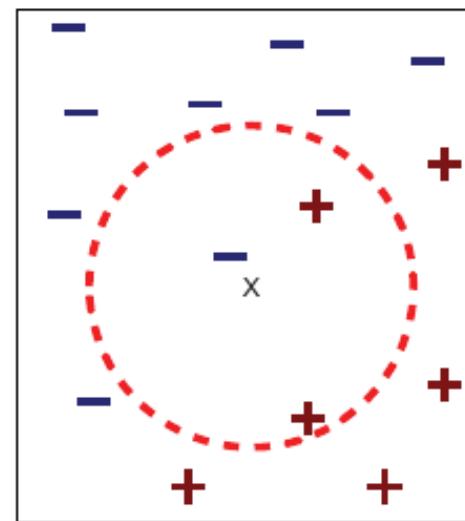
- Classify using the majority vote of the k closest training points



(a) 1-nearest neighbor



(b) 2-nearest neighbor



(c) 3-nearest neighbor

Abordagem k-nn (k nearest neighbors) Aprendizagem por instâncias

- K-NN algorithm does not explicitly compute decision boundaries. The boundaries between distinct classes form a subset of the Voronoi diagram of the training data.

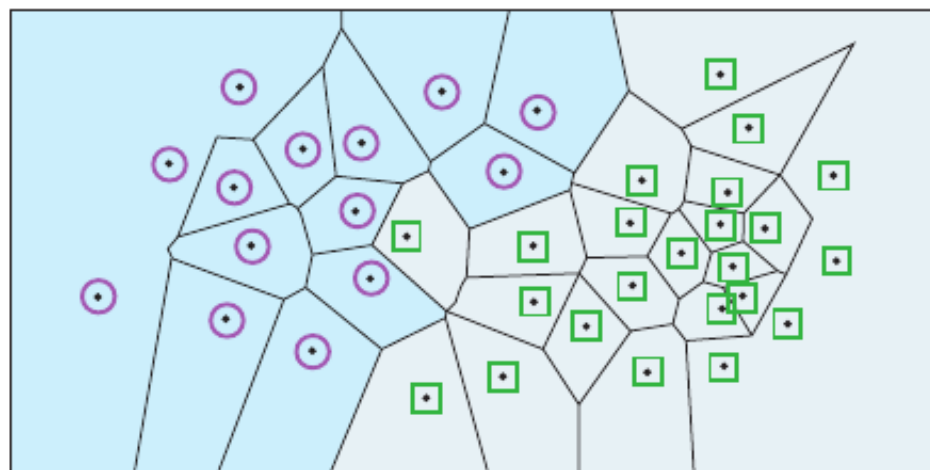


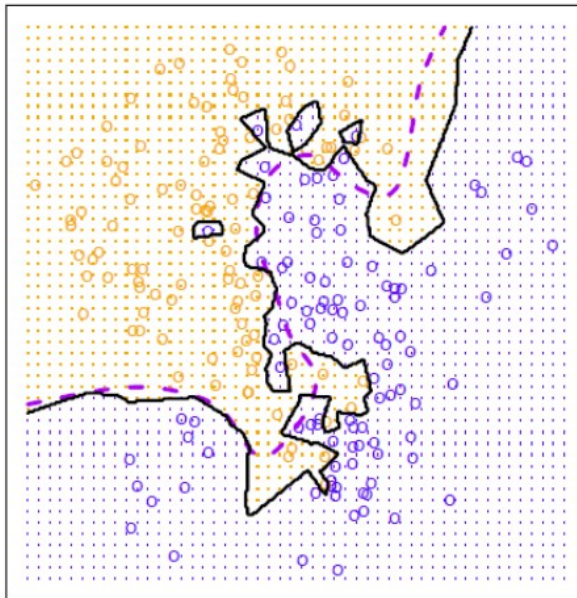
Image by MIT OpenCourseWare.

Each line segment is equidistant to neighboring points.

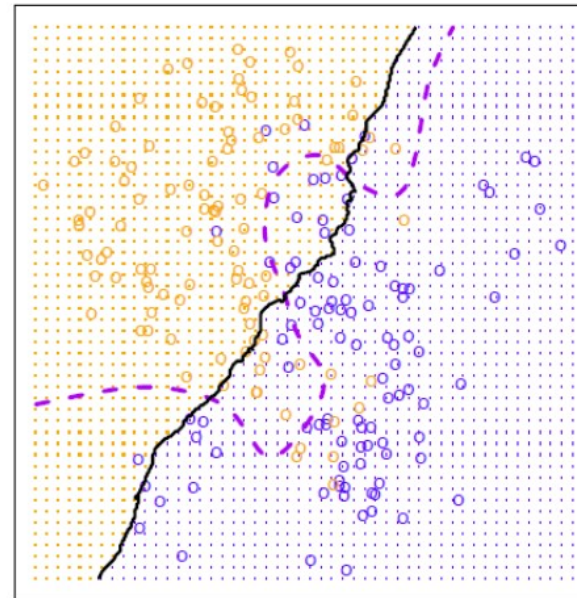
Abordagem k-nn (k nearest neighbors) Aprendizagem por instâncias

- Quanto menor for k, mais flexível será o resultado.

KNN: K=1



KNN: K=100



Abordagem k-nn (k nearest neighbors)

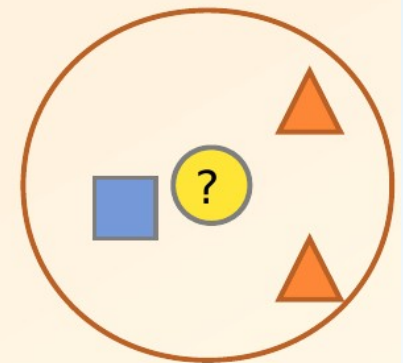
Aprendizagem por instâncias

- Ideia básica:
 - A regra de classificação k-nn é atribuir a uma amostra teste o rótulo da maioria de suas k amostras de treinamento mais próximos.
 - Na prática k é usualmente escolhido como ímpar para evitar empates.

Abordagem k-nn (k nearest neighbors)

Aprendizagem por instâncias

- Requer 3 coisas:
 1. Espaço de Atributos(Dados de treinamento)
 2. Métrica de distância
 - para computar distância entre registros
 3. Um valor de k
 - o número de vizinhos mais próximos para comparar classes
- Para classificar um registro desconhecido:
 - Computar distância dos registros de treinamento
 - Identificar k vizinhos mais próximos
 - Usar rótulos das classes dos vizinhos para determinar classe do registro desconhecido



Abordagem k-nn (k nearest neighbors)

Aprendizagem por instâncias

- Dado um exemplo q para ser classificado
 - Sendo x_1, \dots, x_k os k exemplos de treinamentos em $Tr_exemplos$ mais próximos a q
 - Retornar
$$\sum_{i=1}^k \delta(v, f(x_i))$$
 - Onde V é o conjunto finito de valores de classe e $\delta(a, b) = 1$ if $a = b$, e 0 caso contrário
 - Intuitivamente o algoritmo k-nn atribui a cada nova instância a classe majoritária entre os k vizinhos mais próximos

Abordagem k-nn (k nearest neighbors)

Aprendizagem por instâncias

- **Métricas de distância comuns:**

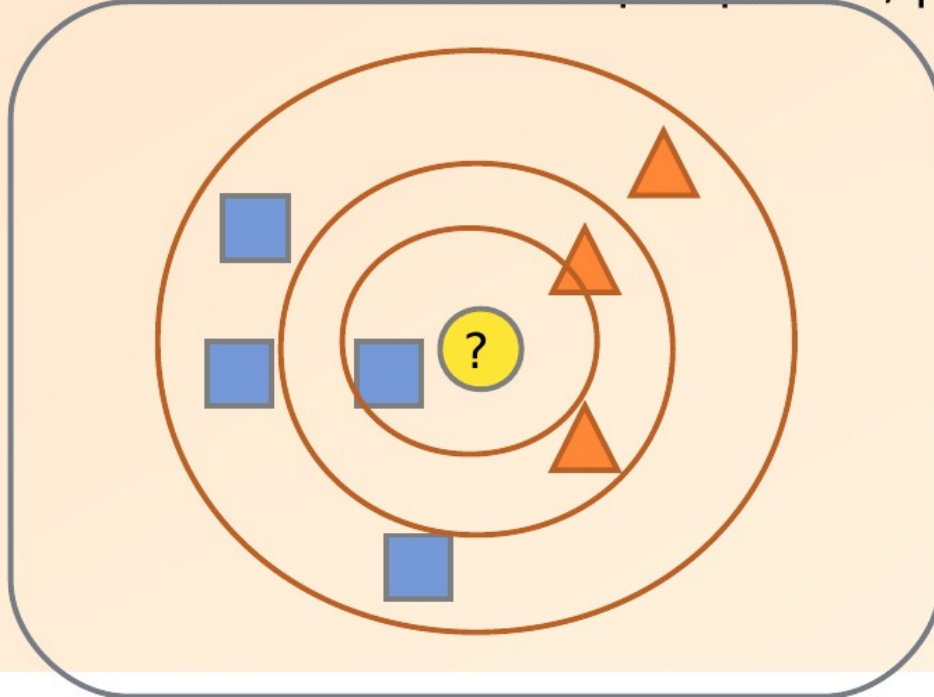
- Distância Euclideana (distribuição contínua)
 - $d(p,q) = \sqrt{\sum (p_i - q_i)^2}$
- Distância de Hamming (métrica com sobreposição)
 - bat (distância = 1) toned (distância = 3)
 - cat roses
- Métrica Discreta (métrica booleana)
 - Se $x = y$ então $d(x,y) = 0$. Caso contrário, $d(x,y) = 1$

- **Determinar a classe da lista dos k vizinhos mais próximos**

- Pegar o voto majoritário dos rótulos entre os k-vizinhos
- Fator de ponderação
 - $w = 1/d$ (interpolação generalizada linear) ou $1/d^2$

Abordagem k-nn (k nearest neighbors) Aprendizagem por instâncias

- Escolhendo o valor de k :
 - Se k for muito pequeno, sensível a ruído
 - Se k for muito grande, vizinhança pode incluir pontos de outras classes
 - Escolher um valor ímpar para k , para eliminar empates



- $k = 1$:
 - Pertence à classe quadrado
- $k = 3$:
 - Pertence à classe triângulo
- $k = 7$:
 - Pertence à classe quadrado

Decisão pelo voto da maioria;

Abordagem k-nn (k nearest neighbors)

Aprendizagem por instâncias

- K-nn funciona em muitos problemas práticos e tolerante a ruídos
- Podem existir muitos atributos irrelevantes
- distância/métrica deve ser adequada

Abordagem k-nn (k nearest neighbors) Aprendizagem por instâncias

- Qual a melhor **distância** usar?
- Qual o melhor valor de **k**?
- As amostras (dados) devem ser analisadas observando **viés** (*bias*) e **variância**

Abordagem k-nn (k nearest neighbors)

Aprendizagem por instâncias

- De forma prática

- Dividir (aleatoriamente) o conjunto de dados em **treinamento** e **teste** (e.g. 70%, 30%)

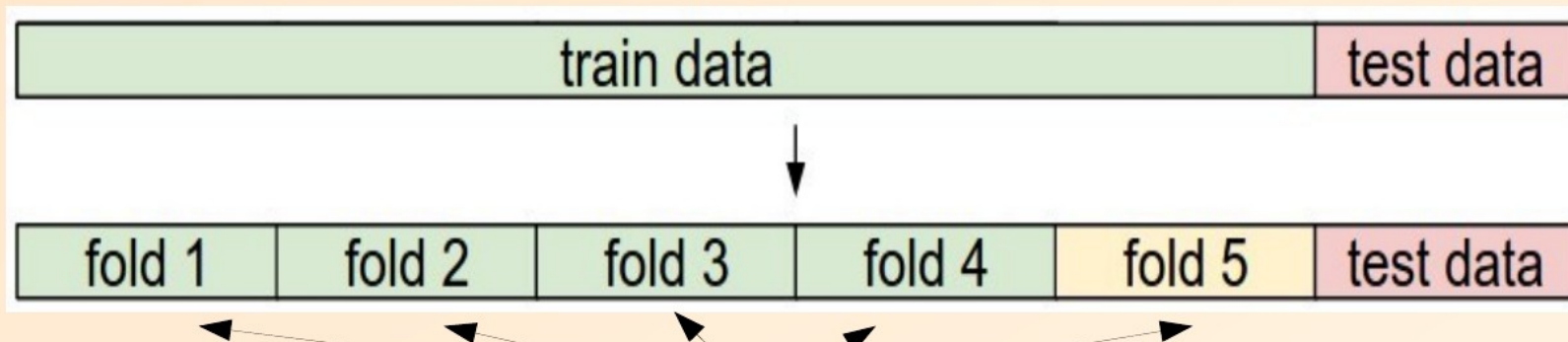


- Depois dividir (também aleatoriamente) a parte de treinamento em um número de partições (e.g. **5** ou 10)

Abordagem k-nn (k nearest neighbors)

Aprendizagem por instâncias

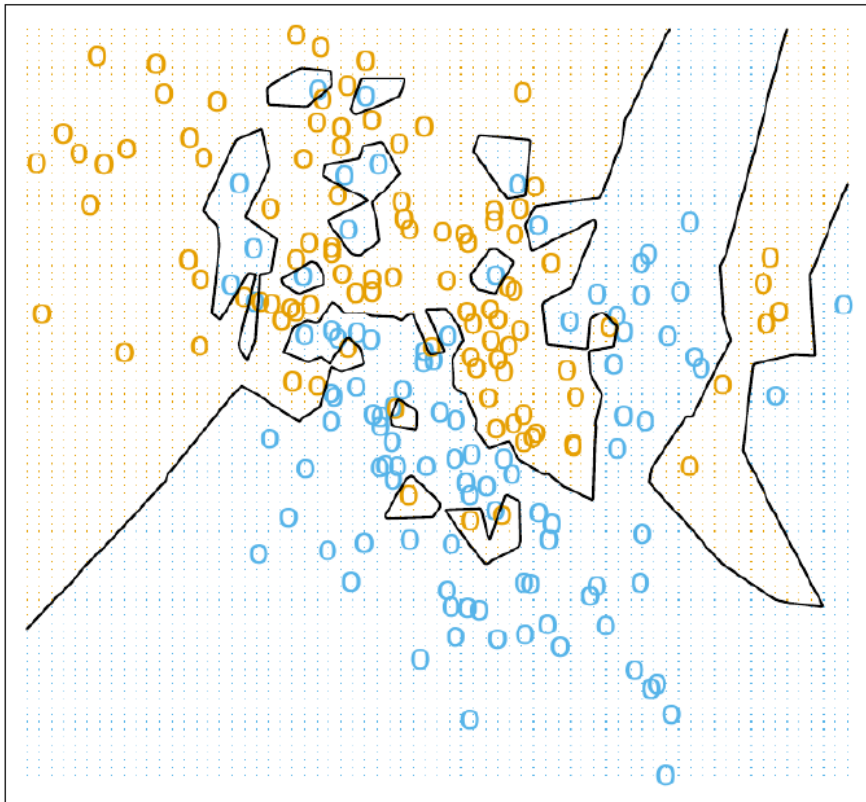
- Escolha uma dessas partições para ajustar/validar esses parâmetros (k e distância)



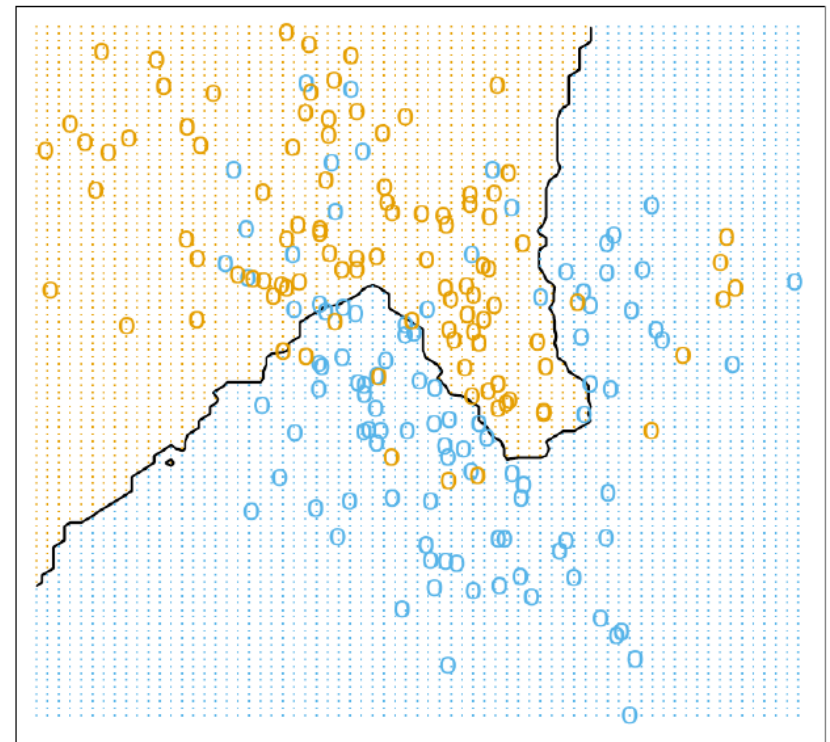
- Use validação cruzada (*cross-validation*) trocando a partição e depois fazendo a média dos resultados no treinamento

Abordagem k-nn (k nearest neighbors) Aprendizagem por instâncias

1-Nearest Neighbor Classifier



15-Nearest Neighbor Classifier



(James et al., 2017)

Abordagem k-nn (k nearest neighbors)

Aprendizagem por instâncias

- Modelo de ajuste por k-nn

```
model = sklearn.neighbors.KNeighborsRegressor(n_neighbors=3)
```

Número de vizinhos



Usando biblioteca Scikit-Learn

Abordagem k-nn (k nearest neighbors)

Aprendizagem por instâncias

Romance, Ação? Quais parâmetros?

Table 2.1 Movies with the number of kicks and number of kisses shown for each movie, along with our assessment of the movie type

Movie title	# of kicks	# of kisses	Type of movie
<i>California Man</i>	3	104	Romance
<i>He's Not Really into Dudes</i>	2	100	Romance
<i>Beautiful Woman</i>	1	81	Romance
<i>Kevin Longblade</i>	101	10	Action
<i>Robo Slayer 3000</i>	99	5	Action
<i>Amped II</i>	98	2	Action

(Harrington, 2012)

Abordagem k-nn (k nearest neighbors)

Aprendizagem por instâncias

Romance, Ação? Quais parâmetros?

Table 2.1 Movies with the number of kicks and number of kisses shown for each movie, along with our assessment of the movie type

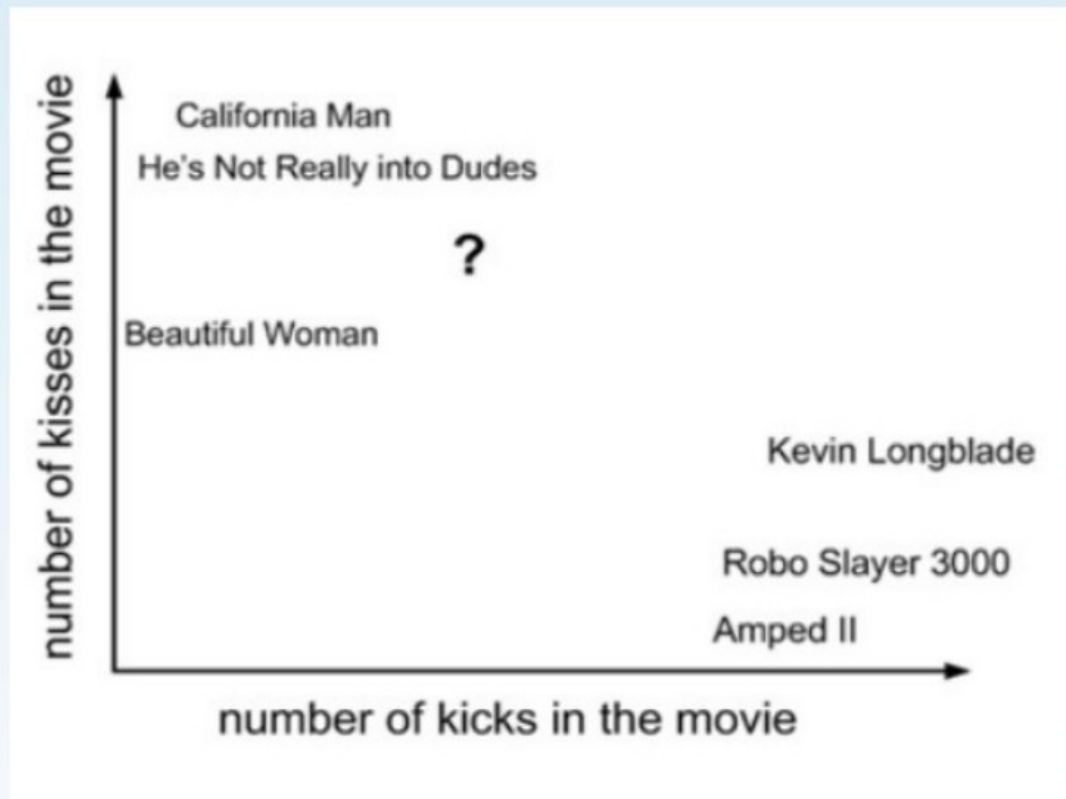
Movie title	# of kicks	# of kisses	Type of movie
<i>California Man</i>	3	104	Romance
<i>He's Not Really into Dudes</i>	2	100	Romance
<i>Beautiful Woman</i>	1	81	Romance
<i>Kevin Longblade</i>	101	10	Action
<i>Robo Slayer 3000</i>	99	5	Action
<i>Amped II</i>	98	2	Action
?	18	90	Unknown

(Harrington, 2012)

Abordagem k-nn (k nearest neighbors)

Aprendizagem por instâncias

Romance, Ação? Quais parâmetros?



Plotar as variáveis

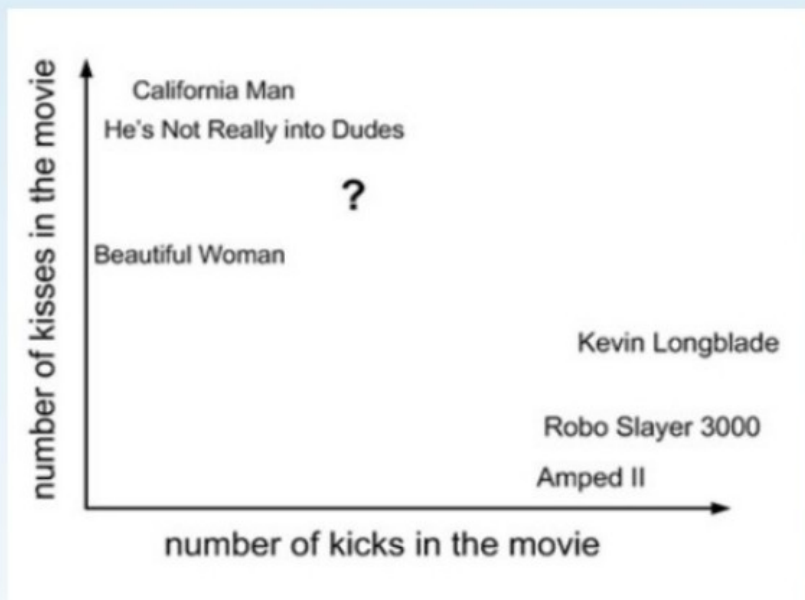
(Harrington, 2012)

Abordagem k-nn (k nearest neighbors) Aprendizagem por instâncias

Romance, Ação?

Quais parâmetros?

Distâncias entre os pontos



Movie title	Distance to movie “?”
California Man	20.5
He's Not Really into Dudes	18.7
Beautiful Woman	19.2
Kevin Longblade	115.3
Robo Slayer 3000	117.4
Amped II	118.9

knn, k=3

(Harrington, 2012)

Abordagem k-nn (k nearest neighbors)

Aprendizagem por instâncias

Pros/Cons to K-NN

Pros:

- Simple and powerful. No need for tuning complex parameters to build a model.
- No training involved (“lazy”). New training examples can be added easily.

Abordagem k-nn (k nearest neighbors)

Aprendizagem por instâncias

Pros/Cons to K-NN

Cons:

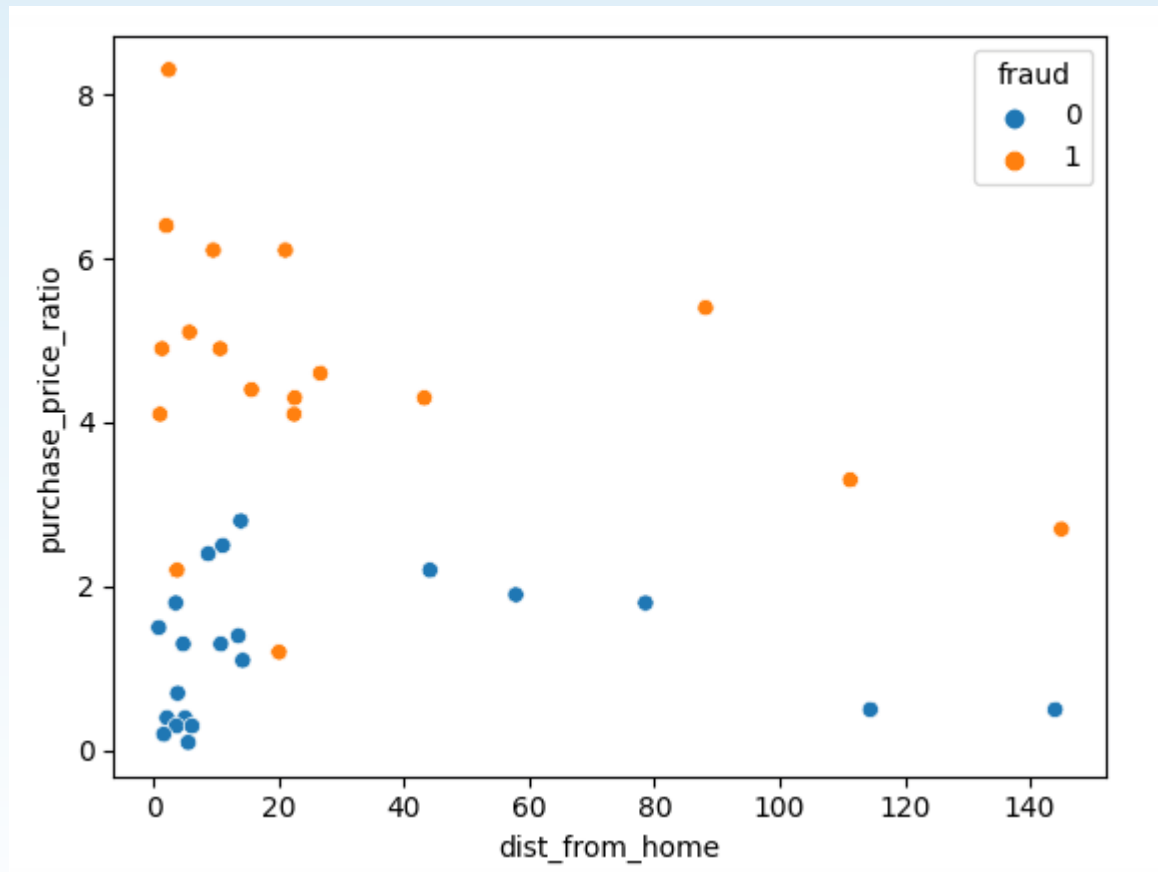
- **Expensive and slow:** $O(md)$, $m = \# \text{ examples}$, $d = \# \text{ dimensions}$
 - To determine the nearest neighbor of a new point x , must compute the distance to all m training examples. Runtime performance is slow, but can be improved.
 - Pre-sort training examples into fast data structures
 - Compute only an approximate distance
 - Remove redundant data (condensing)

Exemplo de projeto k-nn em python

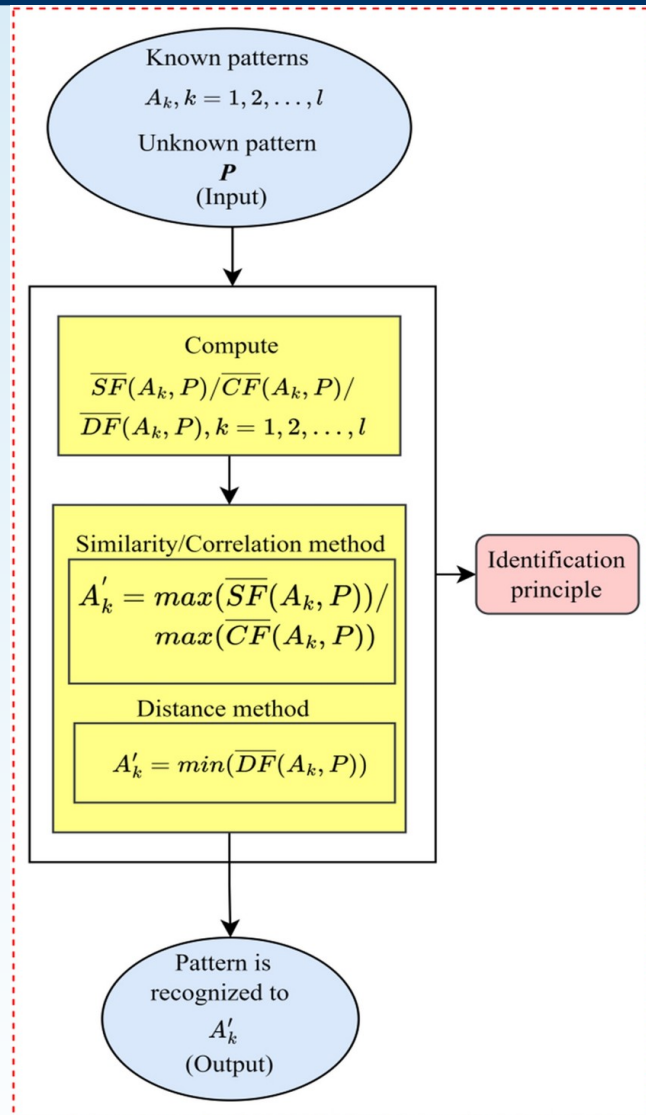
- **Classificar possibilidade de fraude em transações**
 - <https://www.datacamp.com/tutorial/k-nearest-neighbor-classification-scikit-learn>
- **Duas variáveis, 39 observações**
- **dist_from_home: The distance between the user's home location and where the transaction was made.**
- **purchase_price_ratio: the ratio between the price of the item purchased in this transaction to the median purchase price of that user.**
-

Exemplo de projeto k-nn em python

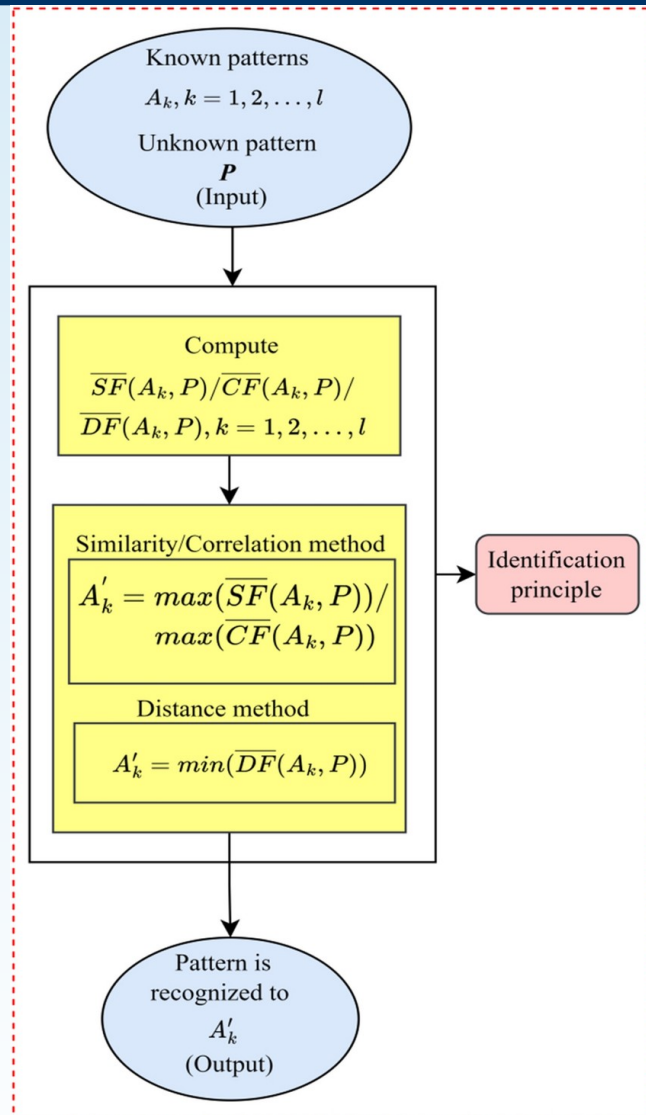
- Classificar possibilidade de fraude em transações
- <https://www.datacamp.com/tutorial/k-nearest-neighbor-classification-scikit-learn>



Similaridade e distâncias em Reconhecimento/aprendizagem de padrões

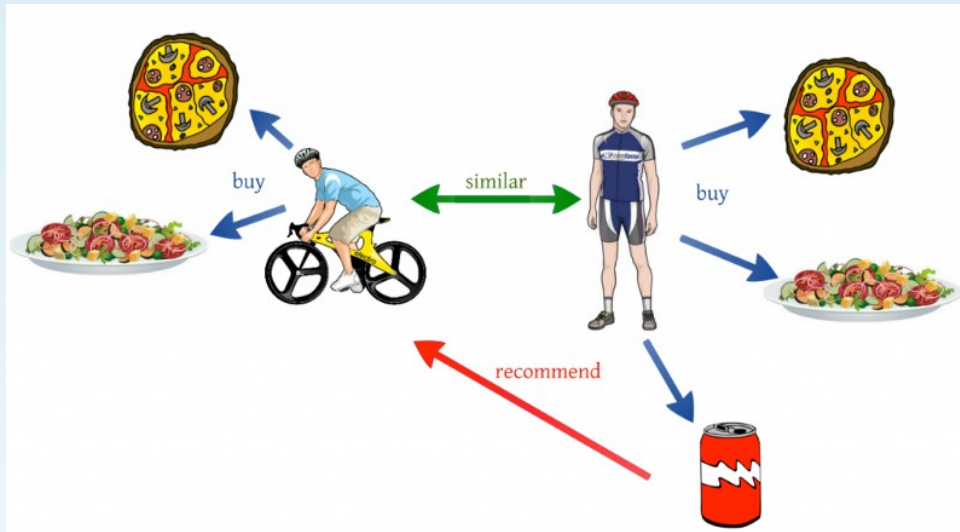


Similaridade e distâncias em Reconhecimento/aprendizagem de padrões



Para analisar se um novo dado/objeto/vetor é igual/semelhante/ (o quanto?) a outro é necessária uma função/norma de similaridade/distância

Similaridade e distâncias em Reconhecimento/aprendizagem de padrões



**Para identificar,
recomendar, filtrar...**

Exemplos de funções de similaridade

MINKOWSKI DISTANCE

$$d(p, q) = \left(\sum_{i=1}^n |p_i - q_i|^C \right)^{\frac{1}{C}}$$

When $C=1$
Minkowski
is Manhattan
distance.

When $C=2$
Minkowski
is Euclidean
distance

ith element of
vector p

ith element
of vector q

Chris Albon

Exemplos de funções de similaridade

MANHATTAN DISTANCE

$$\sum_{i=1}^n |p_i - q_i|$$

Element of a vector
Element of a vector

Imagine you are in Manhattan and can only travel North, South, East, and West.

ChrisAlbon

L1NORM

(Manhattan Norm)

Also Called
"Taxicab Norm"

$$\|x\|_1 = \sum_{i=1}^n |x_i|$$



ChrisAlbon

Exemplos de funções de similaridade

L2 NORM

(Euclidean Norm)

$$\|x\|_2 = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2}$$

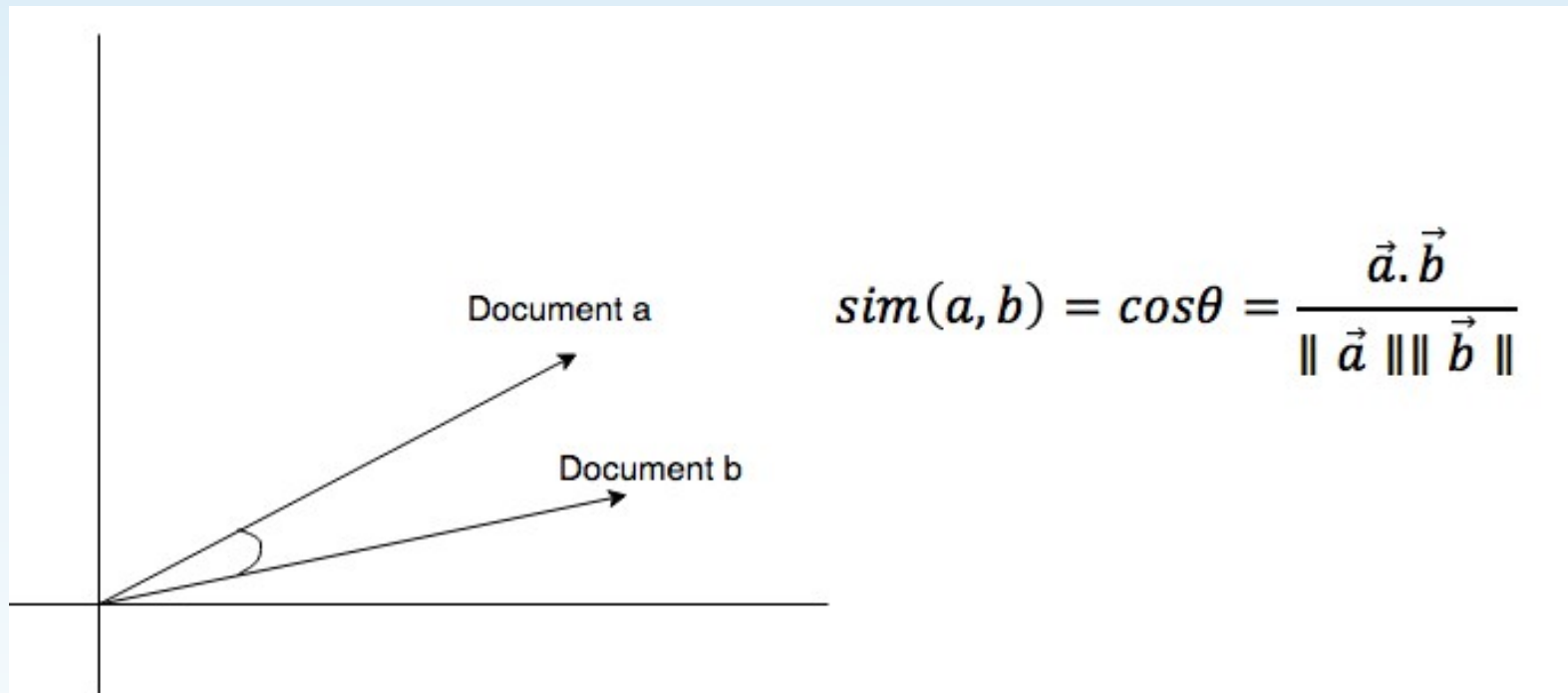
L2 norm is used in many places in machine learning such as normalizing observations and regularization.

↳ Like in NLP.

↳ Example: Ridge Regression.

Chris Albon

Exemplos de funções de similaridade



Exemplos de funções de similaridade

CORRELATION

(Pearson's R)

$$\text{Cor}(X, Y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}}$$

Handwritten annotations above the formula:
- "value of x_i " points to x_i
- "mean of x " points to \bar{x}
- "value of y_i " points to y_i
- "mean of y " points to \bar{y}

Ranges between -1.0 and +1.0. The closer to 0.0 the less linear dependence between variables.

BY CHRIS ALBON

Exemplos de funções de similaridade

Hamming Distance

$$D_H = \sum_{i=1}^k |x_i - y_i|$$

$$x = y \Rightarrow D = 0$$

$$x \neq y \Rightarrow D = 1$$

X	Y	Distance
Male	Male	0
Male	Female	1

HAMMING LOSS

$$L_{\text{Hamming}} = \frac{\sum_{i=1}^n I(\hat{y}_i \neq y_i)}{n}$$

Predicted class \hat{y}_i (green arrow)
true class y_i (blue arrow)
 n (red arrow)
number of observations

ChrisAlbon

Exemplos de funções de similaridade

S. Gupta et al.

Expert Systems With Applications 277 (2025) 127235

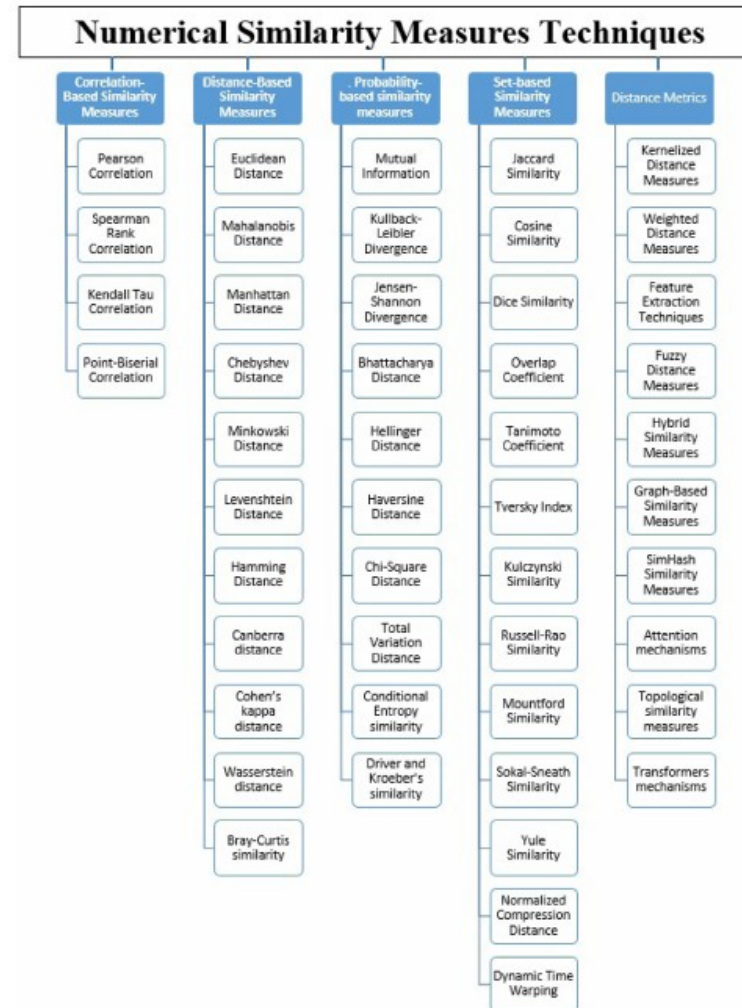


Fig. 2. Categorizes of numerical similarity measure techniques.

Referências Bibliográficas

- Russell, S. & Norvig, P. *Artificial Intelligence: a modern approach* (4th ed), Pearson, 2020. (Capítulo 19)
- Alpaydin, E. *Introduction to Machine Learning*, MIT Press, 2010. (Capítulo 2)
- James, G.; Witten, D.; Hastie, T. & Tibshirani, R. *An Introduction to Statistical Learning with applications in R*, Springer, 2017. (Capítulo 8)
- Mitchell, T. *Machine Learning*. McGraw Hill, 1997.