

Machine Learning com Python

Prof. Luciano Galdino

APRENDIZAGEM BASEADA EM INSTÂNCIAS (KNN)

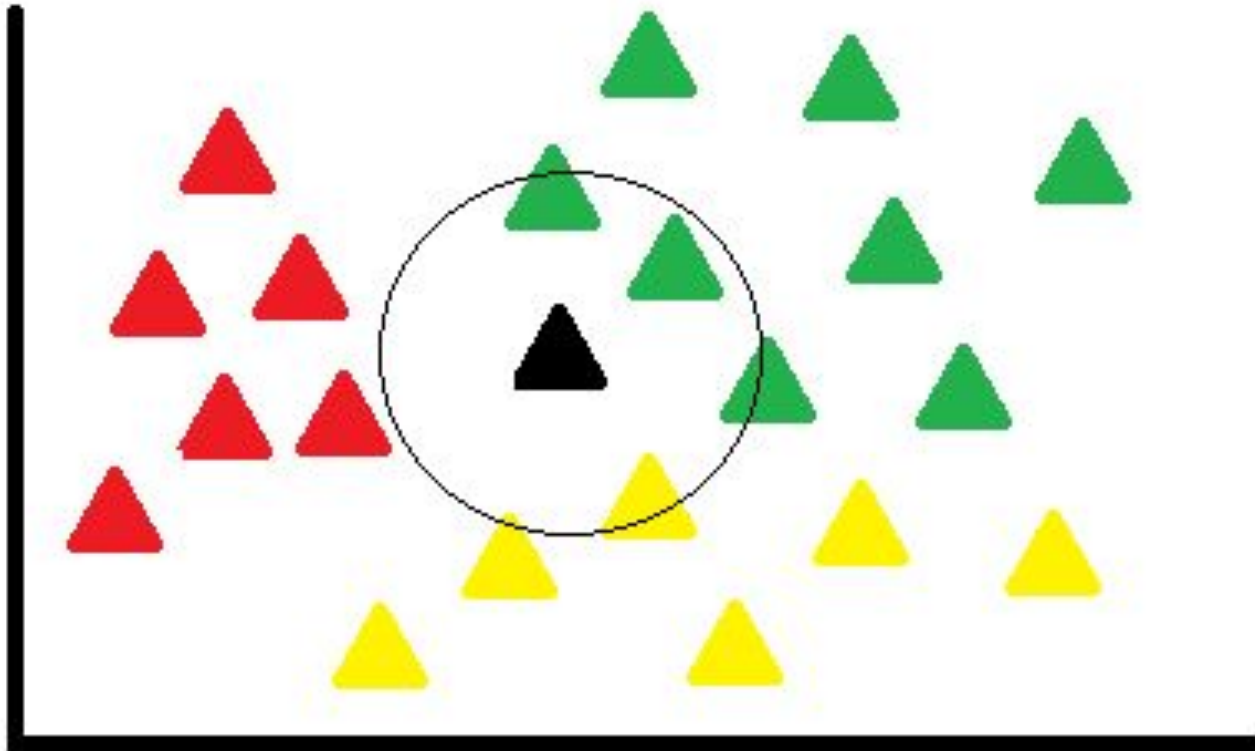
Algoritmo de classificação em aprendizagem supervisionada.

KNN é a sigla de K Nearest Neighbors (K vizinhos mais próximos).

O KNN realiza classificação de instâncias (dados) em classes (grupos de dados semelhantes).

Não possui um modelo matemático, apenas classifica uma instância através de cálculos de distâncias.

Em problemas de classificação é recomendado utilizar número de classes ímpares para evitar empates durante o processo.



[https://www.researchgate.net/figure/A-k-nearest-neighbor-KNN-classifier-KNN-is-explained-as-follo-
ws_fig1_318096864](https://www.researchgate.net/figure/A-k-nearest-neighbor-KNN-classifier-KNN-is-explained-as-follo-
ws_fig1_318096864)

Metrics intended for real-valued vector spaces:

identifier	class name	args	distance function
"euclidean"	EuclideanDistance	•	$\text{sqrt}(\text{sum}((x - y)^2))$
"manhattan"	ManhattanDistance	•	$\text{sum}(x - y)$
"chebyshev"	ChebyshevDistance	•	$\text{max}(x - y)$
"minkowski"	MinkowskiDistance	p, w	$\text{sum}(w * x - y ^p)^{(1/p)}$
"wminkowski"	WMinkowskiDistance	p, w	$\text{sum}(w * (x - y) ^p)^{(1/p)}$
"seuclidean"	SEuclideanDistance	V	$\text{sqrt}(\text{sum}((x - y)^2 / V))$
"mahalanobis"	MahalanobisDistance	V or VI	$\text{sqrt}((x - y)' V^{-1} (x - y))$

Vantagens

Fácil implementação.

Fácil entendimento.

Excelente desempenho em situações de dados com características complexas.

Poucos parâmetros para ajustar.

Desvantagens

Alto custo computacional.

Parâmetro k é ajustado na tentativa e erro.

Necessita transformar dados categóricos em numéricos.