

CÁLCULO DO KNN

A distância euclidiana entre dois objetos p e q é obtida da seguinte forma:

$$d(p, q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$$

Em que n é o número de atributos da base.

No exemplo utilizado na aula, temos o seguinte objeto para identificar qual a linguagem de programação. Vamos chamar esse objeto de obj_{novo} .

departamento	nível	tempo (anos)	linguagem
4	3	2	?

A seguinte base de dados foi usada para o treinamento do KNN. A coluna # serve apenas para referência do número da linha. Ela não faz parte da base de dados.

#	departamento	nível	tempo (anos)	linguagem
1	1	1	1	R
2	2	3	6	Rubi
3	4	2	3	Python
4	2	1	2	Java
5	3	2	4	Java
6	2	2	2	PHP
7	1	2	3	Python
8	4	2	2	Python
9	2	3	10	PHP
10	4	1	1	R
11	1	2	3	Python
12	1	2	3	R
13	4	3	3	Python

Podemos então calcular a distância.

$$d(obj_{novo}, obj_1) = \sqrt{(4-1)^2 + (3-1)^2 + (2-1)^2} = \sqrt{3^2 + 2^2 + 1^2} = \sqrt{14} \cong 3.74$$

IMPACTA

$$d(obj_{novo}, obj_2) = \sqrt{(4-2)^2 + (3-3)^2 + (2-6)^2} = \sqrt{2^2 + 0^2 + 4^2} = \sqrt{20} \cong 4,47$$

$$d(obj_{novo}, obj_3) = \sqrt{(4-4)^2 + (3-4)^2 + (2-3)^2} = \sqrt{0^2 + 1^2 + 1^2} = \sqrt{2} \cong 1,41$$

$$d(obj_{novo}, obj_4) = \sqrt{(4-2)^2 + (3-1)^2 + (2-2)^2} = \sqrt{2^2 + 2^2 + 0^2} = \sqrt{8} \cong 2,83$$

$$d(obj_{novo}, obj_5) = \sqrt{(4-3)^2 + (3-2)^2 + (2-4)^2} = \sqrt{1^2 + 1^2 + 2^2} = \sqrt{6} \cong 2,45$$

$$d(obj_{novo}, obj_6) = \sqrt{(4-2)^2 + (3-2)^2 + (2-2)^2} = \sqrt{2^2 + 1^2 + 0^2} = \sqrt{5} \cong 2,24$$

$$d(obj_{novo}, obj_7) = \sqrt{(4-1)^2 + (3-2)^2 + (2-3)^2} = \sqrt{3^2 + 1^2 + 1^2} = \sqrt{11} \cong 3,31$$

$$d(obj_{novo}, obj_8) = \sqrt{(4-4)^2 + (3-2)^2 + (2-4)^2} = \sqrt{0^2 + 1^2 + 0^2} = \sqrt{1} = 1$$

$$d(obj_{novo}, obj_9) = \sqrt{(4-2)^2 + (3-3)^2 + (2-10)^2} = \sqrt{2^2 + 0^2 + 8^2} = \sqrt{68} = 8,25$$

$$d(obj_{novo}, obj_{10}) = \sqrt{(4-4)^2 + (3-1)^2 + (2-1)^2} = \sqrt{0^2 + 2^2 + 1^2} = \sqrt{5} = 2,24$$

$$d(obj_{novo}, obj_{11}) = \sqrt{(4-1)^2 + (3-2)^2 + (2-3)^2} = \sqrt{3^2 + 1^2 + 1^2} = \sqrt{11} \cong 3,31$$

$$d(obj_{novo}, obj_{12}) = \sqrt{(4-1)^2 + (3-2)^2 + (2-3)^2} = \sqrt{3^2 + 1^2 + 1^2} = \sqrt{11} \cong 3,31$$

$$d(obj_{novo}, obj_{13}) = \sqrt{(4-4)^2 + (3-3)^2 + (2-3)^2} = \sqrt{0^2 + 0^2 + 1^2} = \sqrt{1} = 1$$

Ao final, temos as seguintes distâncias. Os vizinhos mais próximos são: 4, 8 e 13. Com isso, o KNN define que a linguagem do funcionário *obj_novo* é Python.

#	linguagem	Distância em relação a <i>obj_novo</i>
1	R	3,74
2	Rubi	4,47
3	Python	1,41
4	Java	2,83
5	Java	2,45
6	PHP	2,24
7	Python	3,32
8	Python	1,00
9	PHP	8,25
10	R	2,24
11	Python	3,32
12	R	3,32
13	Python	1,00