

04

## Matemática do Davies-Bouldin

### Transcrição

Temos a fórmula de Davies-Bouldin, vamos compreender o que ela significa e quais são seus valores. Começaremos por "R".

#### Índice Davies-Bouldin

$$DB = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} R_{ij}$$

"R" evoca outra fórmula para ser calculado, e basicamente nos entrega uma medida de similaridade entre dois clusters, "i" e "j".

#### Índice Davies-Bouldin

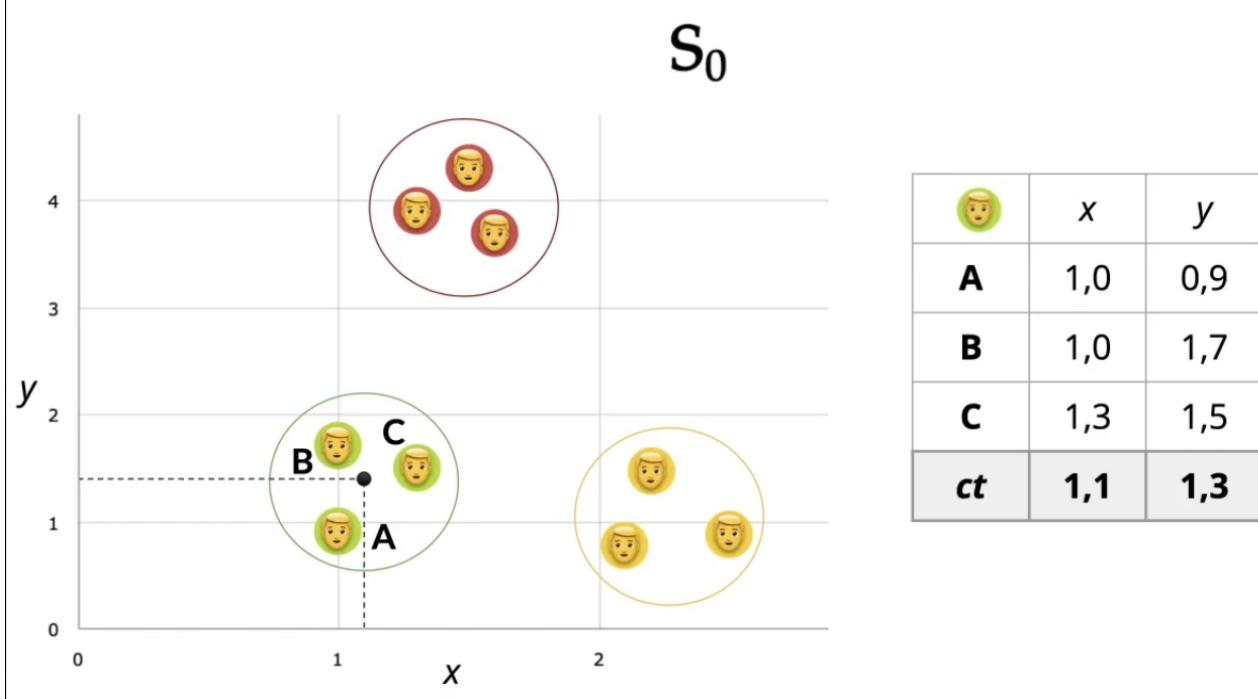
$$R_{ij} = \frac{s_i + s_j}{d_{ij}}$$

"s" é a similaridade dentro do cluster, isto é, **compactação**. Já o "d" reflete a similaridade entre clusters, a **separação**.

"s" será a distância média entre cada ponto do cluster e seu centroide. Calculamos a distância por meio da fórmula de silhouettes na aula anterior, mas a diferença neste caso é que calcularemos a distância de um ponto até o centro do cluster.

Começaremos a calcular o valor de S para o cluster verde, que contém três elementos: A, B e C com os valores de x e y que já calculamos anteriormente, mas ainda resta o valor centróide.

O centróide nada mais é que a média de todos os pontos dentro do cluster para cada um dos atributos, isto é, para x coletaremos os valores de A, B e C e faremos a média, o mesmo se dá com y.



O próximo é tirarmos a média dos pontos até o centróide. Calcularemos a distância de cada um dos pontos até o centróide, por meio da distância euclidiana.

## Distância Euclidiana

$$d(A, ct) = \sqrt{(x_{ct} - x_A)^2 + (y_{ct} - y_A)^2}$$

$$d(A, ct) = \sqrt{(1,1 - 1,0)^2 + (1,3 - 0,9)^2}$$

$$d(A, ct) = \sqrt{0,17}$$

$$d(A, ct) = 0,41$$

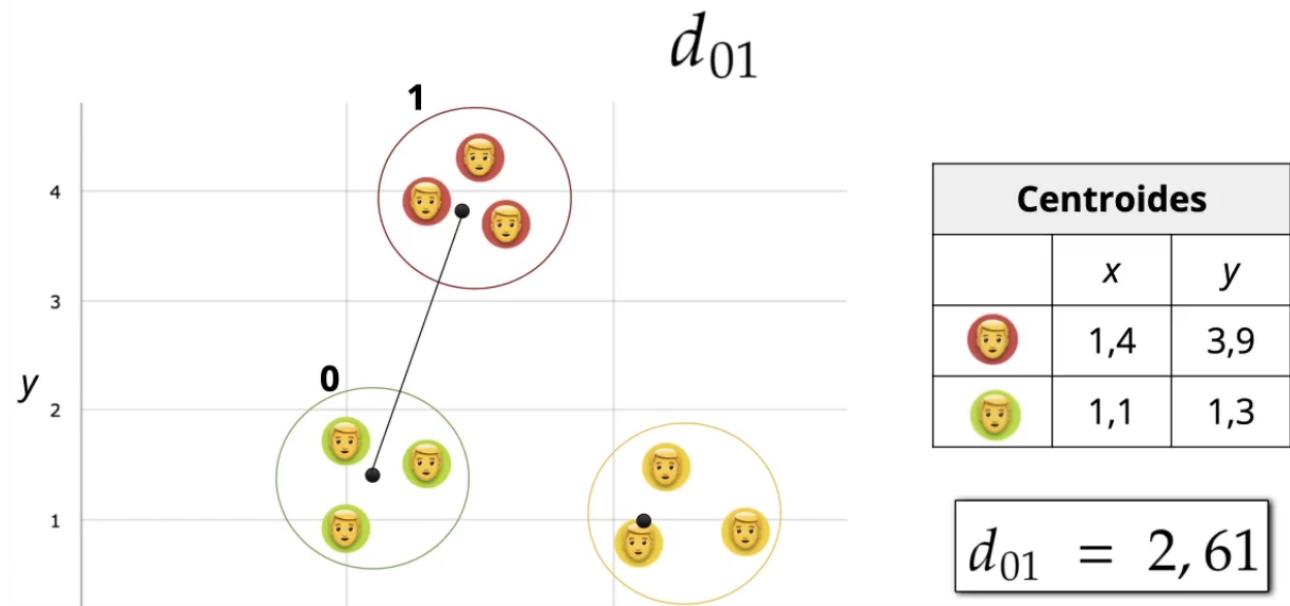
	$x$	$y$
<b>A</b>	1,0	0,9
<b><math>ct</math></b>	<b>1,1</b>	<b>1,3</b>

Faremos o mesmo procedimento para todos os outros pontos a distância de A até o centróide é de 0,41, de B 0,41 e de C 0,28. Ao retirarmos a média, temos o resultado de 0,36, que representa a similaridade dentro de um cluster.

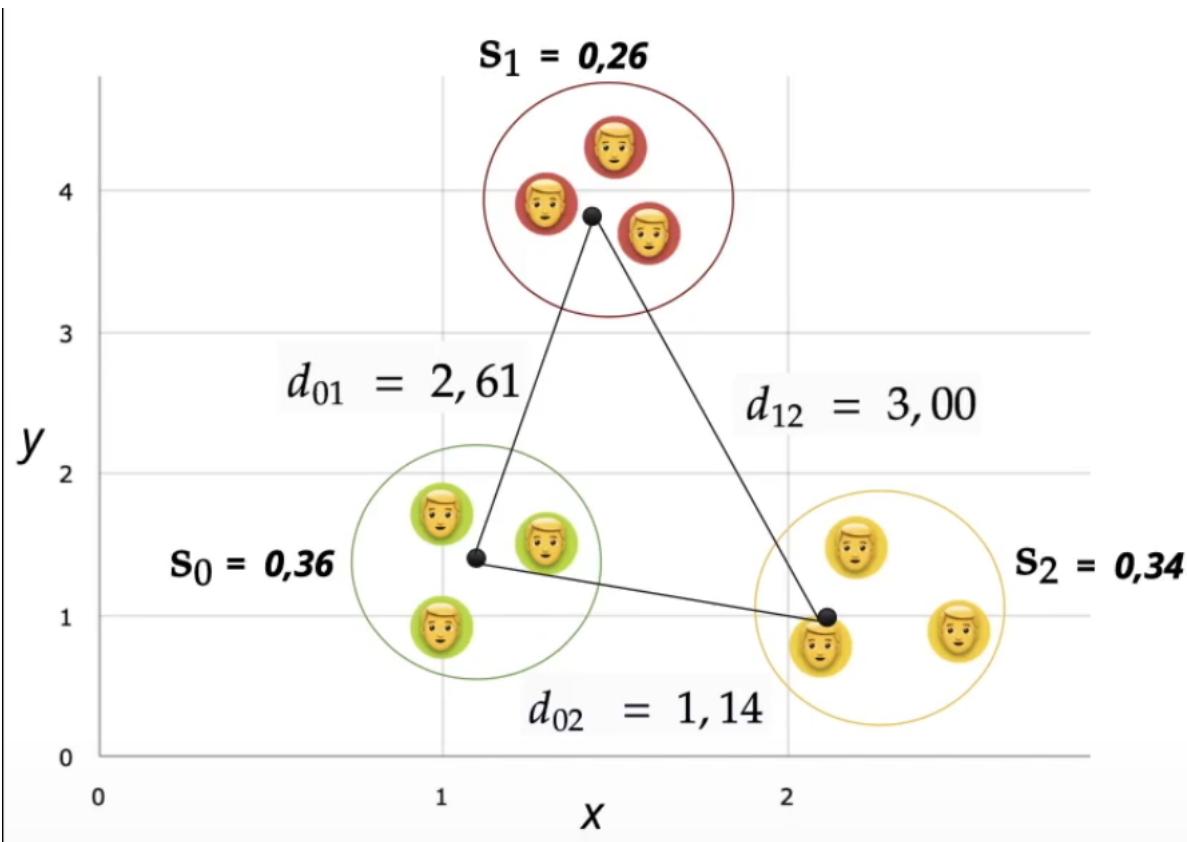
Repetiremos o procedimento com o cluster vermelho e amarelo.



A próxima etapa é descobrir a distância entre os centróides dos clusters. Começaremos com a distância do cluster verde e vermelho. Utilizaremos os valores que adquirimos por meio da distância euclidiana, mas dessa vez baseada unicamente nos centróides. O resultado será de 2,61.



faremos o mesmo procedimento para descobrir a distância entre os outros centróides.



Faremos a conta para obter os resultados de R. Começaremos calculando o valor de R para o cluster 0 e 1. O resultado será 0,23.

$$\frac{s_0 + s_1}{d_{01}} = \frac{0,36 + 0,26}{2,61}$$

$$\frac{0,62}{2,61} \quad \boxed{0,23}$$

Repitiremos o mesmo procedimento para outros clusters. Os valores serão: 0,23; 0,61 e 0,20.

Retornaremos ao índice Davies-Bouldin, e para cada um dos clusters pegaremos o maior valor de R. Para o cluster 0, por exemplo, o valor será 0,61; para o cluster 1 0,23 e para o cluster 2 0,61.

O maior valor reflete o pior resultado, afinal quanto mais próximo de 0, melhor. Para analisarmos a eficiência do nosso cluster, precisamos coletar qual seria seu pior exemplo.

Nosso resultado final será de 0,48.

$$DB = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} R_{ij}$$

$$DB = \mathbf{0,48}$$