

Matemática do Silhouette

Transcrição

Temos a fórmula para calcular o coeficiente que se baseia em dois principais valores: beta e alfa.

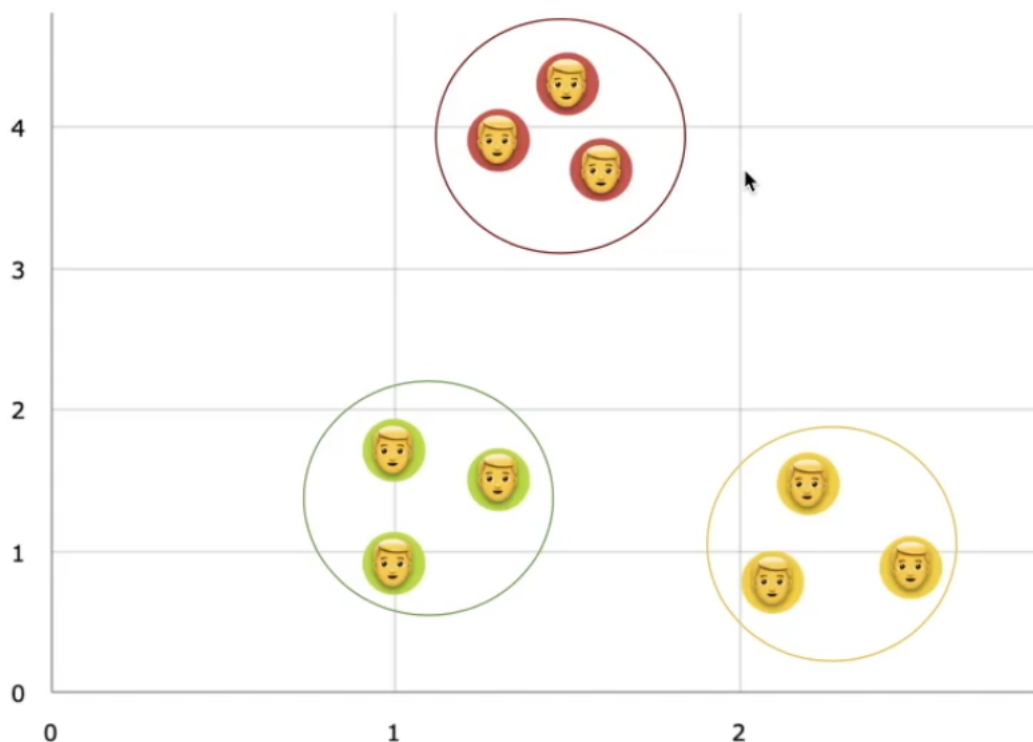
Coeficiente de Silhouette

$$s = \frac{b - a}{\max(a, b)}$$

Começaremos por calcular o valor de alfa, isto é, a distância média entre o ponto e todos os outros pontos do **mesmo cluster**. Vamos verificar a compactação dos elementos.

Utilizaremos o mesmo exemplo dos três clusters, e cada um contém três elementos.

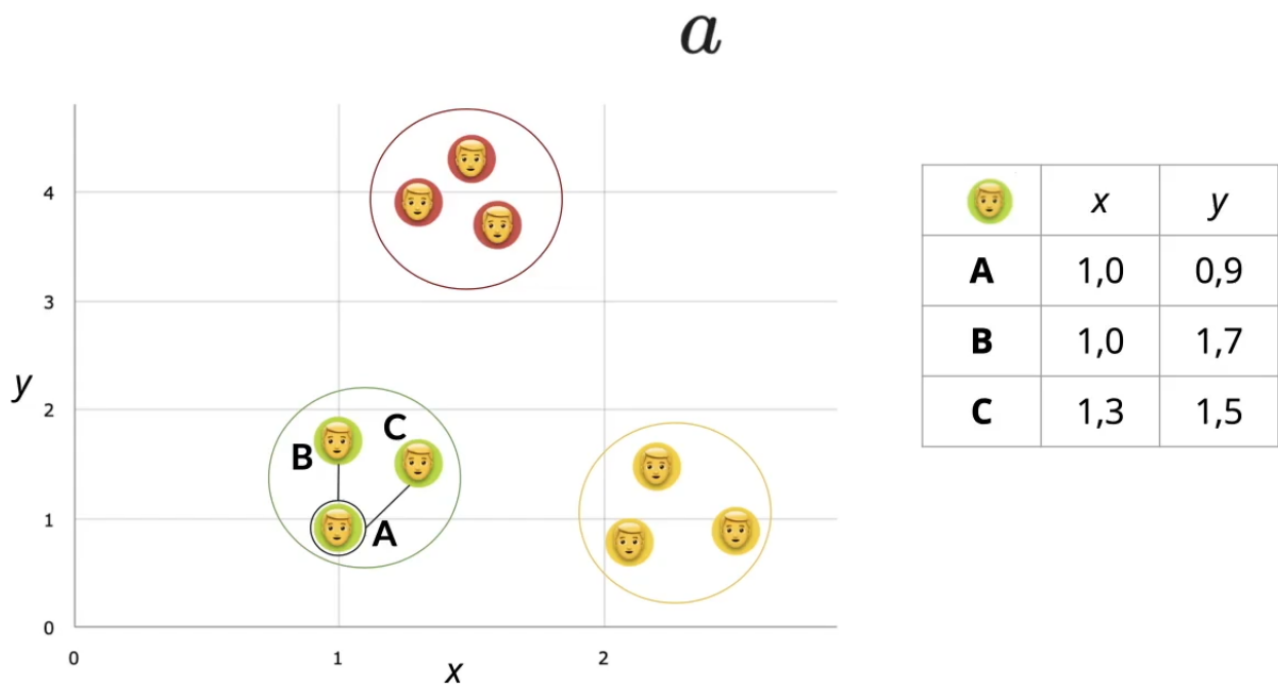
Compactação



Nos basearemos em duas dimensões(x,y), e utilizaremos dois atributos dos 16 que temos disponíveis em nosso dataframe.

Selecionaremos um ponto específico, um cliente, e a partir dele calcularemos a média para outros elementos do cluster. Chamaremos este primeiro ponto de "A".

Para o eixo x utilizaremos o valor de 1,0 e para y 0,9. Feito isso, mediremos a distância do ponto "A" para o ponto "B" e "C"



Estamos utilizando a distância euclidiana, portanto usaremos sua fórmula.

a

Distância Euclidiana

$$d(A, B) = \sqrt{(x_B - x_A)^2 + (y_B - y_A)^2}$$

$$d(A, B) = \sqrt{(1 - 1)^2 + (1,7 - 0,9)^2}$$

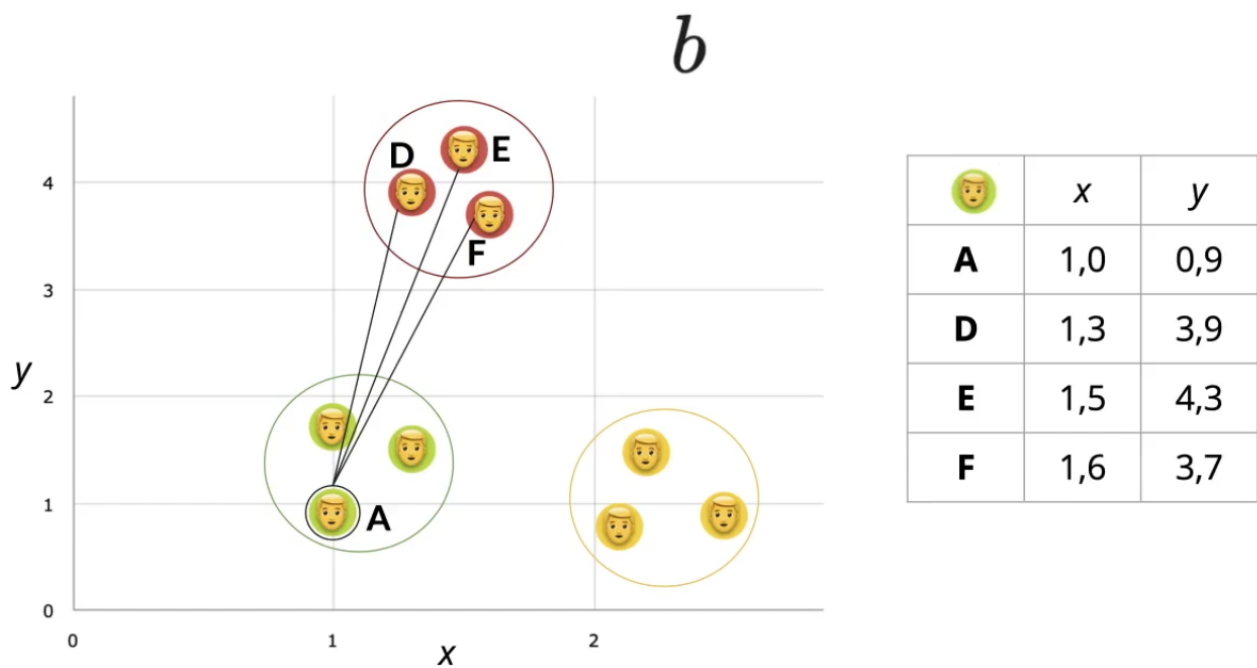
	x	y
A	1,0	0,9
B	1,0	1,7
C	1,3	1,5

O resultado dos cálculos será de 0,8. Essa é a distância Euclidiana entre A e B. Faremos o mesmo procedimento para C. A distância entre o ponto A e C é 1,12.

Feito isso, obteremos a média entre os dois valores finais, o resultado é 0,96. Voltaremos a fórmula do coeficiente de silhuete que apresentamos no início da aula e inseriremos esse valor na variável alfa.

Nosso próximo passo é calcular o valor de beta, que é a distância média entre o ponto e todos os outros pontos do cluster mais próximo.

Para sabermos qual é o cluster mais próximo, deveremos tirar a média entre os pontos para os dois clusters e coletar o menor valor. Faremos o mesmo procedimento que realizamos anteriormente, mas com pontos diferentes.



chegaremos ao resultado final de 3,1 do ponto A dos pontos do cluster vermelho. Para o cluster amarelo o resultado será de 1,31. A menor distância é o segundo cluster, como podemos verificar.

Coeficiente de Silhouette

$$s = \frac{1,31 - 0,96}{\max(0,96, 1,31)}$$

O resultado final será de 0,26. O resultado sempre está entre 1 e -1, então neste caso temos um bom valor positivo.