

04

Critérios de validação

Transcrição

Geramos nossos clusters, a próxima etapa é de validação deles. Trata-se de uma fase crucial, que definirá se podemos progredir no projeto ou fazer ajustes.

Vamos conhecer algumas métricas de validação existentes, que são divididas basicamente em duas categorias: **métricas externas e internas**.

No caso das métricas externas, precisamos ter labels para realizar cálculos e gerar a validação. Como não temos labels, não as utilizaremos neste projeto. Já as métricas internas independem de labels, ainda que possuam uma estrutura mais complexa.

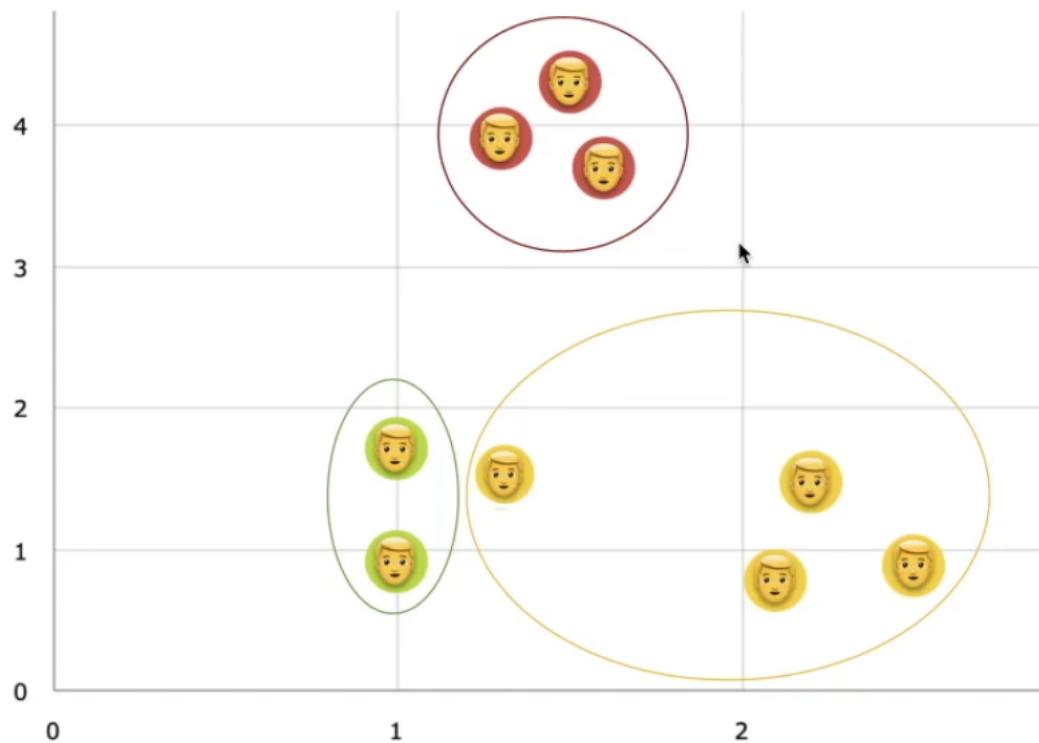
As métricas internas se baseiam em dois critérios de validação, e o primeiro deles é **compactação**, que nos relata quanto próximos estão os pontos em um mesmo cluster. Quanto mais próximos os pontos, mais compactado ele está.

Utilizaremos dois exemplos visuais para compreender melhor essa ideia. No primeiro caso, executamos nosso algoritmo e ele nos gerou três grupos, em que os elementos estão bem compactados, com valores próximos entre si.

Compactação



Já no segundo caso, a compactação não é tão boa para um determinado grupo, tendo um elemento dissidente, mais afastado dos outros valores do cluster.



Isso indica que o segundo resultado é menos eficiente que o primeiro.

O segundo critério de validação é a **separação**, que indica quão bem separados estão os pontos em clusters diferentes. A ideia é justamente o contrário da compactação, quanto mais distantes estão os clusters um dos outros, melhor.

Ao observarmos o primeiro exemplo de compactação, notaremos que os grupos estão bem divididos, apesar dos elementos internos de cada um deles estarem agrupados.

No segundo exemplo, a separação não está boa, afinal o elemento amarelo está muito próximo do verde. Neste caso não temos nem uma boa separação nem uma boa compactação.

Nas próximas aulas, veremos como esses critérios são aplicados na linguagem matemática.