

O número de clusters

Quando estamos um algoritmo de agrupamento como o K-Means, temos que passar o número de grupos que desejamos para segmentar os dados.

Quando passamos um número muito pequeno, temos um problema que dados que não pertencem ao mesmo grupo, podem ficar agrupados no mesmo cluster.

Qual o problema que enfrentamos se o número de clusters for muito alto?

Selecione uma alternativa

- A** Quanto mais grupos, maior a possibilidade de termos dados que pertenceriam a grupos diferentes no mesmo cluster. Isso acontece pois o algoritmo tenta agrupar os dados baseado no número de clusters que passamos
- B** No final, não importa qual o número de grupos que passamos para o algoritmo, pois ele sempre vai otimizar o número de grupos. Isto é, se passarmos 2 como número de clusters e existirem 3 grupos, o algoritmo utilizará 3 como o número certo
- C** Quanto mais grupos, mais segmentamos os dados ficam. Isso pode causar de existirem dados muito próximos - que pertenceriam ao mesmo grupo - fiquem em grupos diferentes. Logo, o agrupamento não estaria correto