



Exercício k-means

No exercício proposto, o algoritmo de separação *k-means*, um algoritmo de divisão em *clusters* (agrupamentos em torno de um ponto) de uma base de dados foi implementado. O algoritmo recebe os dados de entrada e uma quantidade final k de agrupamentos em que pretende-se obter após sua compilação, além de retornar os centros dos *clusters* encontrados e a classificação quanto a qual classe k uma amostra X_i pertence.

Para testar o algoritmo, foram utilizados dados amostrados em torno dos pontos médios $(2, 2)$, $(2, 4)$, $(4, 2)$ e $(4, 4)$, cada vez amostrados com um desvio padrão $\sigma \in \{0.3, 0.5, 0.7\}$. O valor do número de clusters foi também variado entre três possíveis valores, $k \in \{2, 4, 8\}$. Os resultados são mostrados nas figuras abaixo. Em todas as figuras os centróides encontrados ao final do algoritmo estão representados por um símbolo $+$, enquanto os diferentes *clusters* são diferenciados pela cor.

A figura 1 será usada como comparação, uma vez que a sua separação em *clusters* foi satisfatória de acordo com os dados gerados. Nota-se nas figuras seguintes que o valor de k pode ser sobredimensionado, criando-se um *overfitting*, como mostrado na figura 4, em que as classes foram subdivididas de maneira assimétrica. Além disso, o valor de k pode ser subdimensionado, criando-se o *underfitting*, como mostrado na figura 7. Por fim, o valor do desvio padrão σ dos dados amostrados pode atrapalhar o algoritmo *k-means*, tornando difícil a separação dos dados em classes bem definidas. Isto pode ser observado comparando-se as figuras 1 e 3 por exemplo.

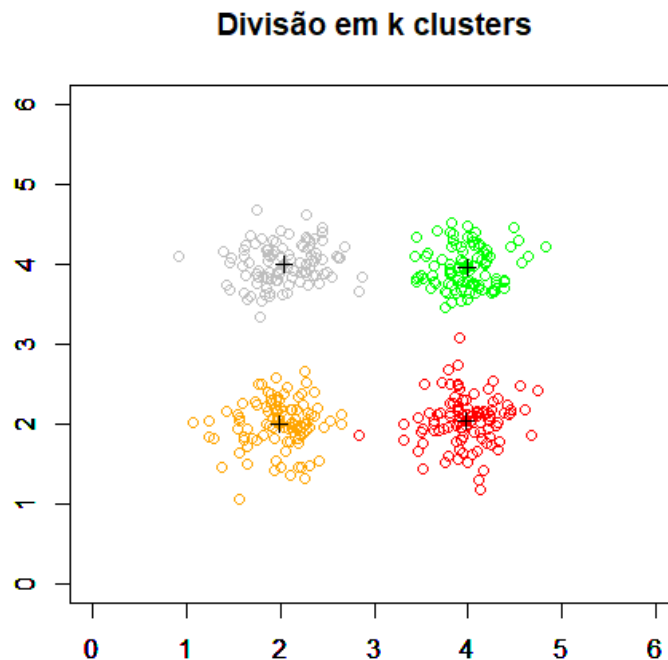


Figura 1: Resultados obtidos para $k = 4$ e $\sigma = 0.3$.

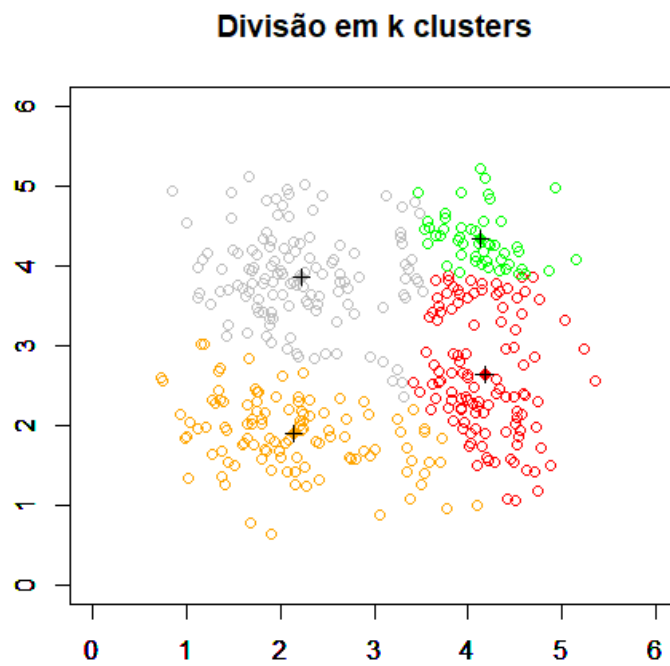


Figura 2: Resultados obtidos para $k = 4$ e $\sigma = 0.5$.

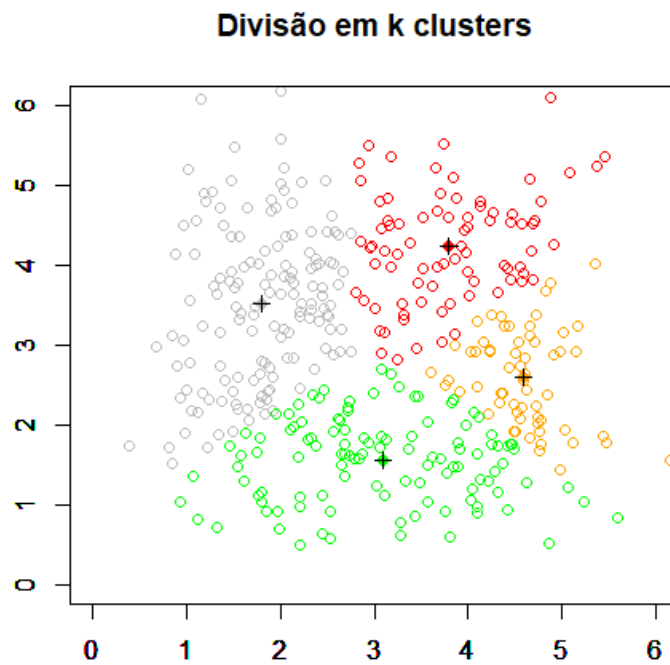


Figura 3: Resultados obtidos para $k = 4$ e $\sigma = 0.7$.

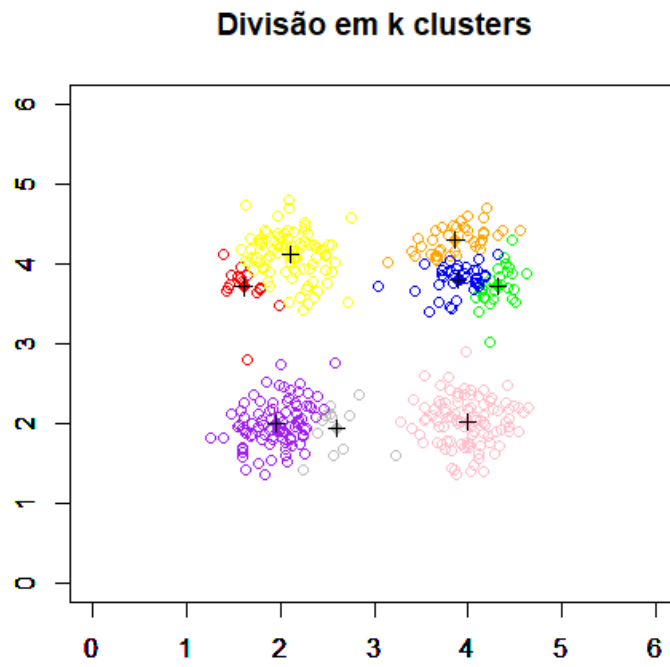


Figura 4: Resultados obtidos para $k = 8$ e $\sigma = 0.3$.

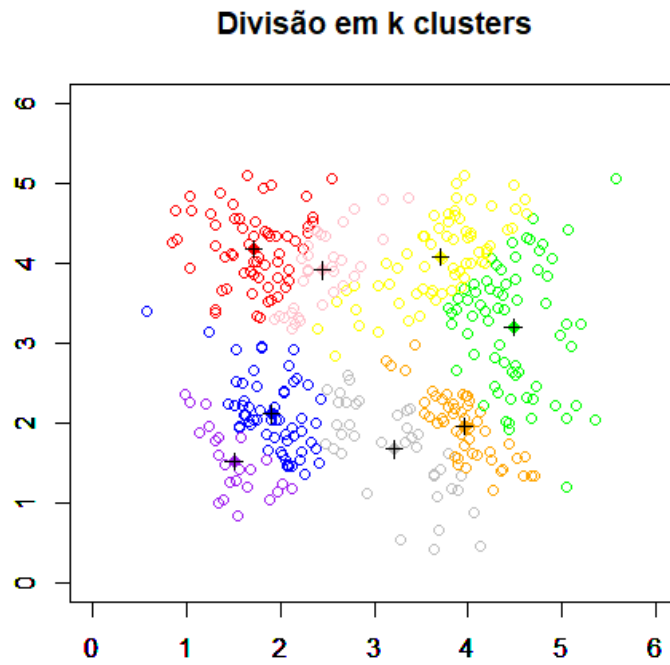


Figura 5: Resultados obtidos para $k = 8$ e $\sigma = 0.5$.

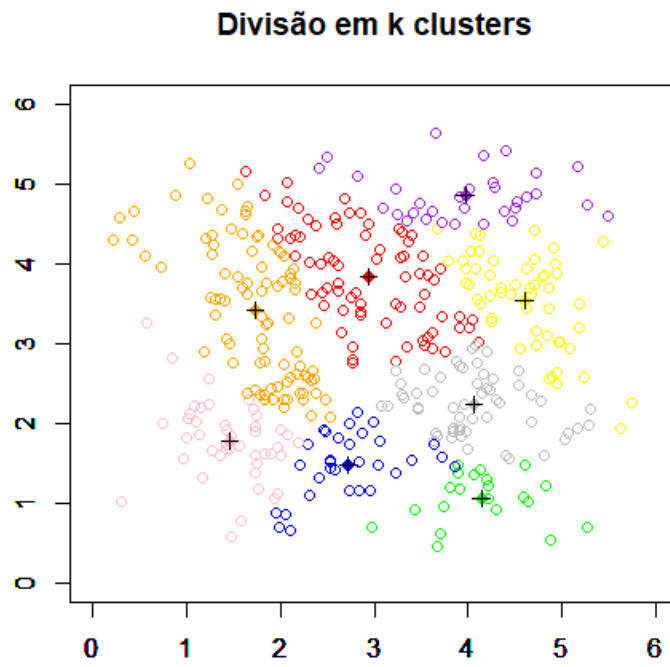


Figura 6: Resultados obtidos para $k = 8$ e $\sigma = 0.7$.

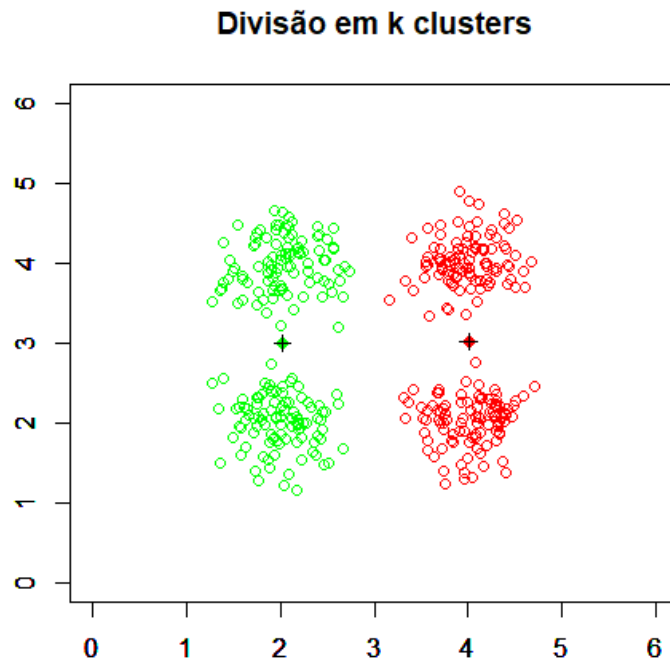


Figura 7: Resultados obtidos para $k = 2$ e $\sigma = 0.3$.

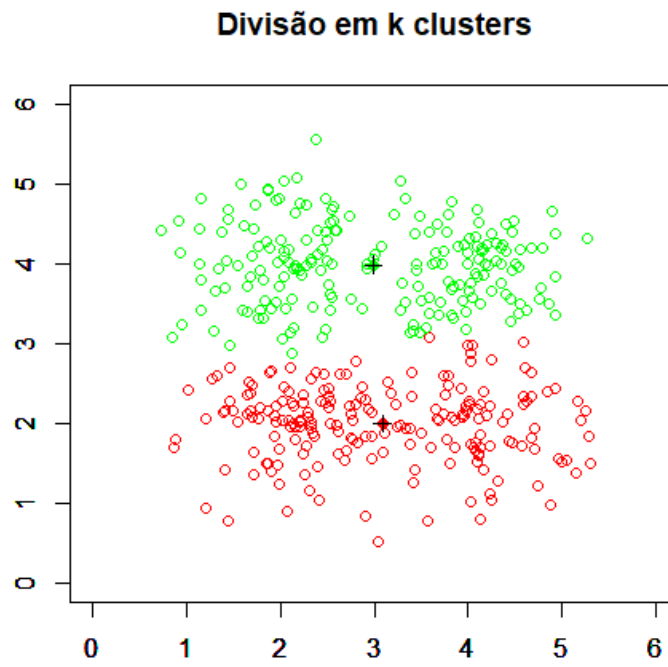


Figura 8: Resultados obtidos para $k = 2$ e $\sigma = 0.5$.

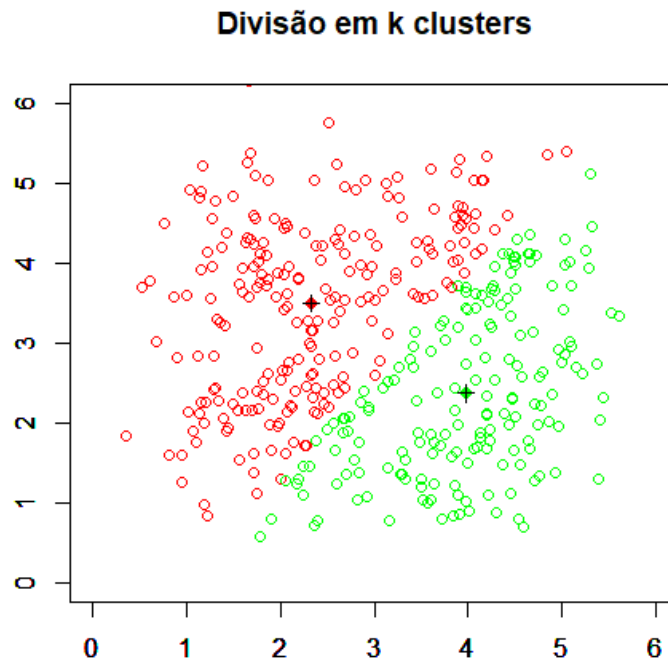


Figura 9: Resultados obtidos para $k = 2$ e $\sigma = 0.7$.