



Exercício k-means 2

No exercício proposto, foram gerados dados em formato de espiral pertencentes a duas classes distintas, conforme mostra a figura 1.

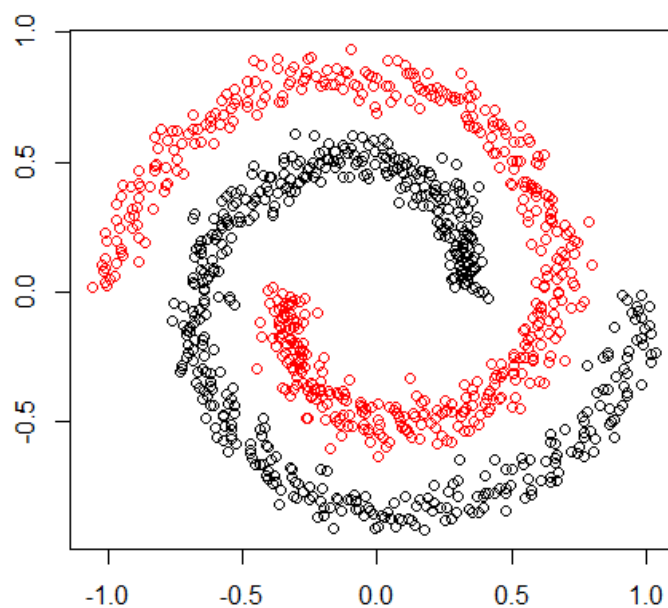


Figura 1: Dados de entrada para o problema.

Foi então decidido em quantos *clusters* seriam divididos os dados utilizando o algoritmo *k-means*. Para isso, o algoritmo foi rodado em *looping* com *k* sendo incrementado em uma unidade a cada iteração. O critério de parada foi de que a média dos valores classes de cada amostra dentro de cada *cluster* não podia ter uma diferença maior que 1% do valor das classes exatas. Por exemplo, um *cluster* que originalmente era da classe 1 deve ter a média da classificação das classes dos dados entre $0.99 < \mu < 1.01$. Com isso, o algoritmo retornou 22 agrupamentos, cuja divisão e centros são mostrados na figura 2.

Por fim, foi calculada a PDF dos agrupamentos encontrados e gerada o gráfico da função de densidade de probabilidade mostrado na figura 3.

É interessante ressaltar que o algoritmo *k-means* implementado pode gerar saídas diferentes de acordo com os centróides iniciais definidos. Com isso, o número mínimo de *clusters*, seguindo o critério das médias das classificações inferiores a 1% estabelecido e ao fazer 10 simulações com os mesmos dados de entrada, obtidos foram valores de *k* entre 19 e 23 grupos, sendo o valor 22 a moda das simulações e, portanto, o valor escolhido como resultado neste trabalho.

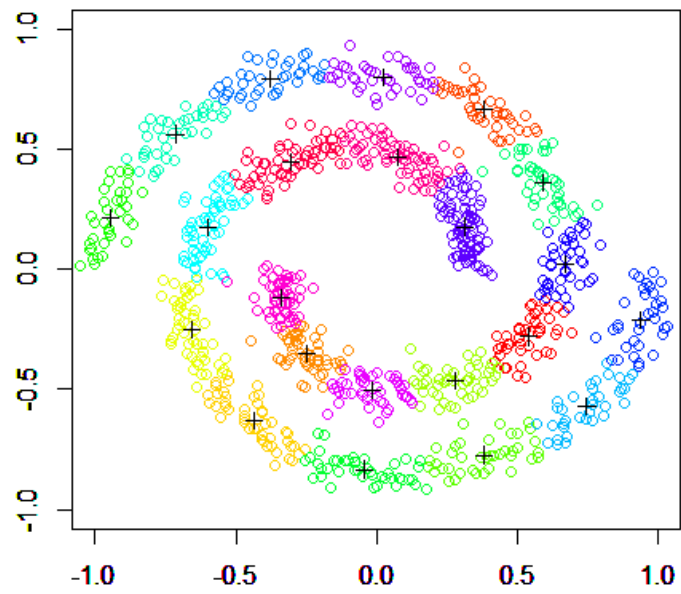


Figura 2: Dados de entrada divididos em $k = 22$ classes.

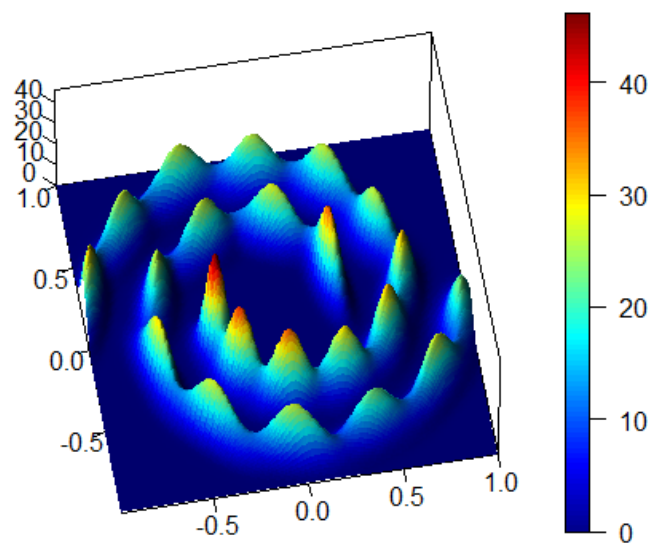


Figura 3: PDF para os dados amostrados.