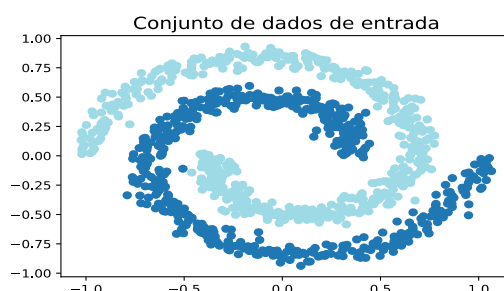

RELATÓRIO EXERCÍCIO K-MEANS SPIRAL

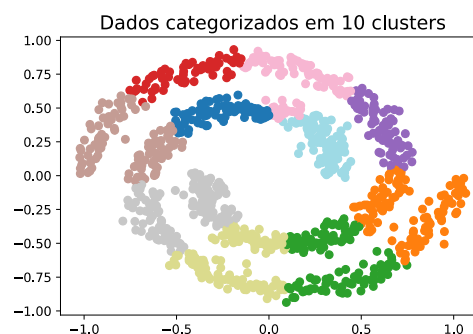
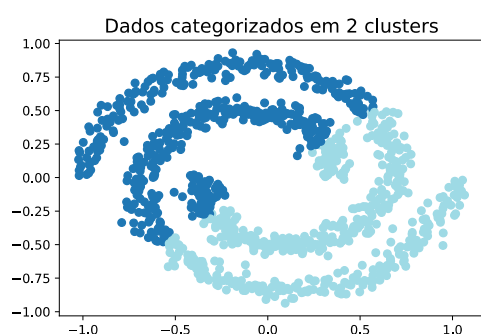
Gabriel Saraiva Espeschit – 2015065541

30 de agosto de 2020

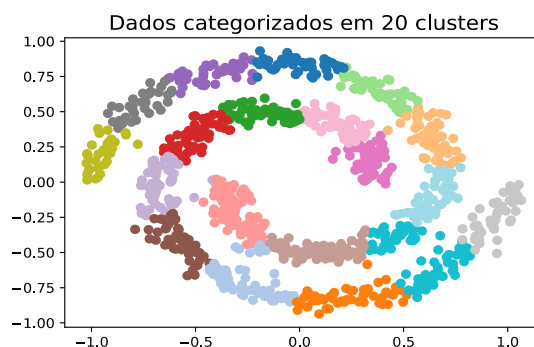
Primeiramente, foi utilizado a biblioteca *pandas* para ler os dados do documento *dataEspiral.txt*. Abaixo pode-se verificar os dados lidos:



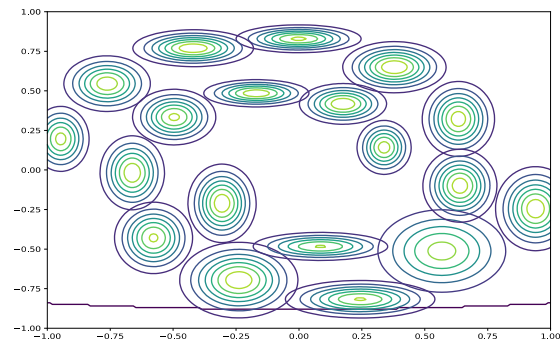
Em seguida se aplicou a função *kmeans* desenvolvida no exercício anterior para os dados de entrada, variando a quantidade de *clusters* até que os *clusters* não contenham dados de 2 classes diferentes.



O valor encontrado para que essa restrição seja satisfeita foi de 20 *clusters*. Para garantir a convergência em tempo hábil do algoritmo nessa situação foi necessário introduzir um parâmetro *max_it* no algoritmo. Esse parâmetro interrompe o algoritmo depois que *max_it* iterações forem percorridas.



Por fim, calculamos a média e o desvio padrão de cada *cluster* e plotamos as funções de densidade de probabilidade.



Com isso podemos averiguar que o método *k-means* é excelente para fazer a classificação de dados que estão claramente separados por um semiplano. No entanto, quando a separação de dados é mais complexa, o método *k-means* dificilmente consegue realizar a separação de dados e outro método deverá ser usado.