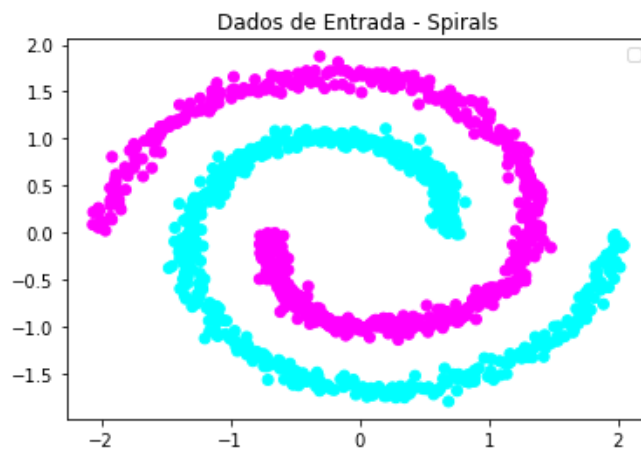


RELATÓRIO KDE

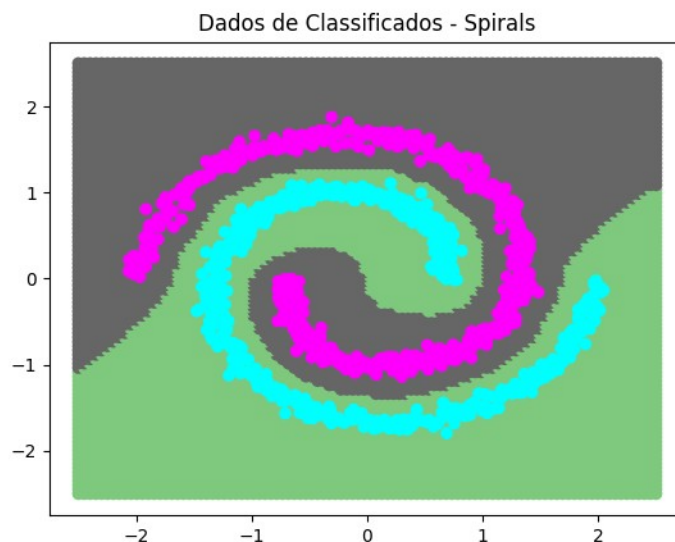
Gabriel Saraiva Espeschit – 2015065541

27 de set de 2020

Para a realização do exercício de KDE, foi utilizado o *dataset* espiral do pacote *mlbench* do R.

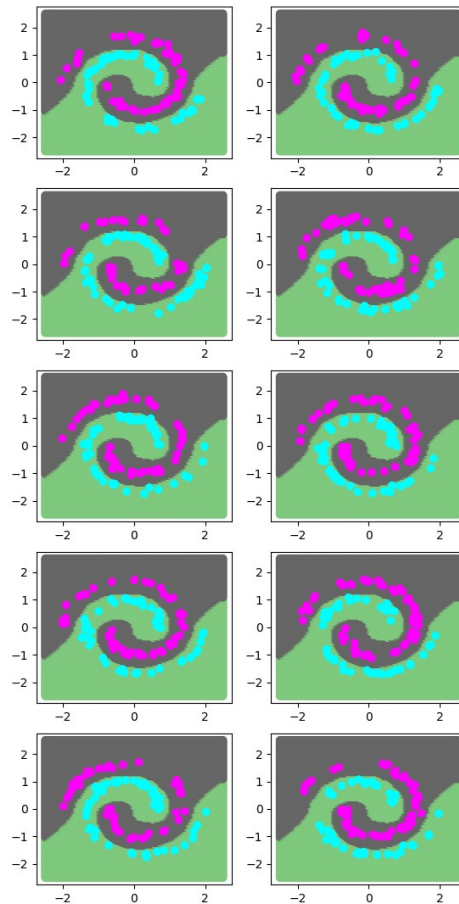


Em seguida, implementou a estimativa por KDE e em seguida implementou a classificação por Bayes. Foi possível gerar um espaço de separação de acordo com essa implementação como se pode ver abaixo quando se passa o parâmetro $h = 0.1$:

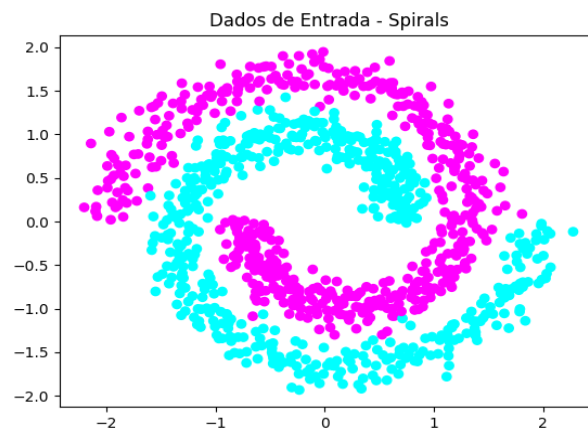


Em seguida dividiu-se os dados em 10 *folds*, para qual foram plotados os dados de teste sob o espaço de classificação gerado.

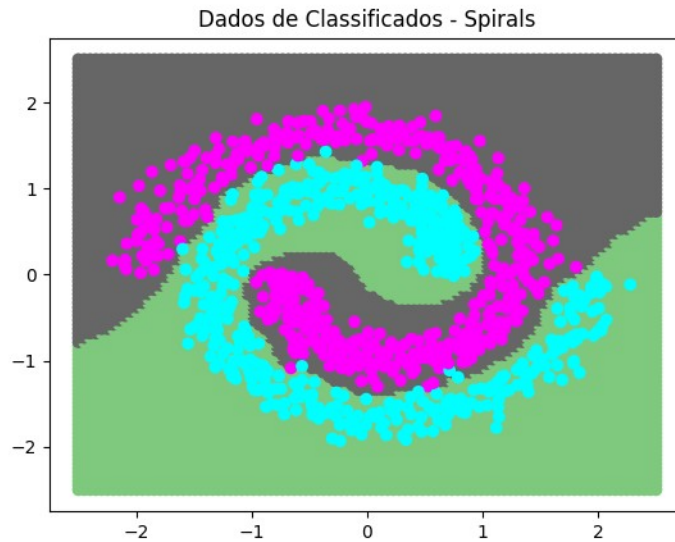
Dados de teste no espaço de separação



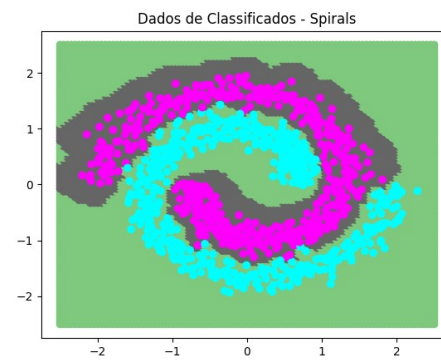
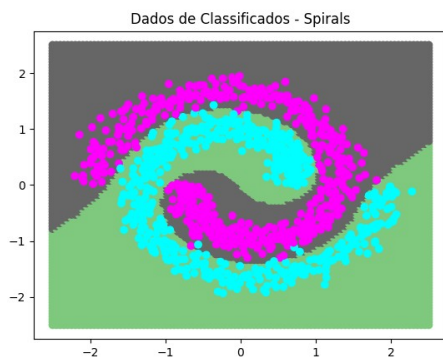
A acurácia média encontrada foi de 100%. Isso significa que o $h = 0.1$ é adequado para resolver esse problema. Sendo assim, aumentou-se o desvio padrão das amostras a fim de gerar um problema mais complicado para resolução e estudar o impacto do h . Os novos dados gerados estão plotados a seguir:



Em seguida, plotou-se os o espaço de separação para esses dados com h também igual a 0.1:



Fez-se a mesma coisa para $h = 1.06\sigma N^{-1/5}$ e $h = 0.01$. Os resultados estão dispostos nas imagens abaixo:



Como pode ser observado, houve uma espécie de *overfitting* para os dados em rosa quando se utilizou o termo h igual a 0.01. Em termos de acurácia quando fazendo o teste e treino usando 10 *folds* obtivemos os seguintes resultados:

Valor de h	Acurácia obtida no 10 <i>folds</i>
0,1	98,9%
$1,06*\sigma*N^{-1/5}$	99,2%
0,01	98,4%

Como podemos ver, realmente ocorreu um *overfitting* quando utilizamos o h como 0,01 e a acurácia obtida pelo modelo foi a menor entre todos os casos testados. O KDE é uma excelente forma de encontrar a separação de dados não linearmente separáveis, no entanto, é importante ficar atento à calibração do h para que os dados não sofram com *over* ou *underfitting*.