4.1

%%

% P error Nºsvm

% 1 46% 100

% 10 0% 95

% 15 21% 93

% 5 16% 91

% 9 0% 96

% 11 2% 99

% 6 0 35

% valores pedidos 4.1

% parecem que quando o valor do polinómio é par vai surgir uma curva de

% separação de classes fechada o valor do erro não varia substancialmente

% quando se aumenta a ordem, sendo que os primeiros são horriveis

% NOW LEGIT

% o número de vectores de suporte para a criação do SVM é sempre

% elevado,sendo que para o caso polinomial vai ser dificil conseguir

% adaptar limites sem usar muito vectores de suporte

% logo não sei o que dizer, isto é aborrecido hi, how are u

%

% o existe uma janela de graus do polinómio que causa que os valor do erro não sejam muito elevados

% seja mais baixo, sendo que neste caso cria

4.2

Quanto menor o sgima vais vetores usa sendo que vai fazer um contorno mais elaborado em torno de cada vetor, sendo utilizados todos os vetores como vetores de suporte pois nestas situações o algoritmo vai tentar considerar cada um dos vetores como individual, fazendo que não exista erro nos vetores de treina no entanto muito pouco flexível mais concretamente ocorre overfitting. à medida que se vai aumentado o

simga consegue-se notar que se vai criando uma área que engloba os datassets , sendo cada vez menos rígida até a um ponto em que começa a englobar pontos pertencentes ao outro (sinal)

Como existiam vários valores em que o erro é o nos escolhemos o valor de

sigma para o qual eram utilizados menos vetores de suporte, formando assim uma zona de decisão menos restritiva, tendo assim possivelmente menor erro de classificação para a inserção de novos valores.

4.3

Sigma = 1

Corresponde a 10 vetores de suporte

4.4

% em vez de fazer uma bolsa individual para cada um dos cluster de dados

% pertencentes à mesma classe criou bubbles em volta de um dos outleirs

%sigma -> vectors

% mas parece que deve ser possivel retirar mais um vector

%01 -> 10

%2 -> 10

%1.5 -> 15

% 2.38 valor mais alto com 10, no entanto não se consegue correr o que vem

Quando foi testada a data com a inserção dos outliers, com o mesmo sigma foi possível detetar que agora vão ser criadas secções pequenas que englobam os outliers ou que irá englobar um conjuntos de dados separando-se dos outros valores .

Devido ao facto da boxcontraint ter o valor **inf** os algoritmo vai fazer com que não seja possível encontrar valores mal classificados logos irá dar uma maior importância ao facto de os valores estarem bem classificados em vez de criar uma boa margem de decisão. Fazendo que as margens possam atingir valores muito baixos, como ocorre neste caso, sendo que os valores actuais de margens são muito inferiores ao que encontramos usando os mesmo parâmetros de classificação com o set de data sem outliers. Devido ao facto de o classificador ter que acertar em todos os dados classificados corretamente vai fazer que agora seja necessária uma adição de vectores de suporte para que tal seja possível. Quando não existia outliers tínhamos 10 vectores de suporte 1 por cada cluster de dados e um adicional para o cluester do mei, agora existem 17 vectores de suporte o aumento vem do facto de serem necessária vários vectores para isolar os outliers existentes.