

Lista 6

Support Vector Machines

Instruções

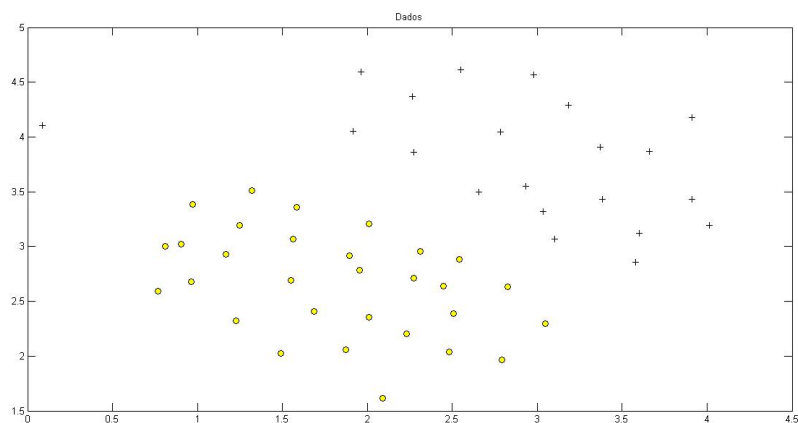
Deverá ser enviado ao professor, um arquivo texto contendo os gráficos, resultados e comentários requeridos em cada item.

1. Classificação usando SVM e ajuste do parâmetro C

- Carregue os dados contidos no arquivo ex6data1.data.

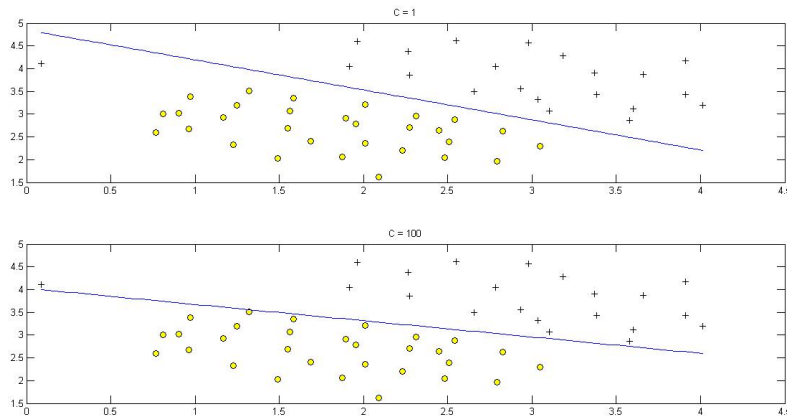
O arquivo contém uma matriz e um vetor de dados. A matriz X é composta de 51 linhas e 2 colunas e representa um conjunto de dados de dimensão 2. O vetor y dá a classe a qual pertence cada vetor. Este exemplo consiste em um problema de duas classes.

Apresentar: Figura com o conjunto de dados



- Utilize a função `svmTrain` para treinar um modelo. Para esta função utilize o Kernel linear. Utilize valores de $C = 1$ e $C = 100$.
- Plote o resultado utilizando a função `visualizeBoundaryLinear`.

Apresentar: Figuras com o conjunto de dados e as superfícies de separação



Comentários: Comente sobre as superfícies de separação obtidas para os dois valores de C

=> O parâmetro C indica o quão próxima estará a superfície de separação dos dados. No caso de $C=100$, notamos que a linha se aproxima mais das duas classes, causando um maior overfitting. No caso de $C=1$, a superfície está numa região mais neutra, quase equidistante das duas classes, sendo então um modelo mais genérico.

- Utilize $C=0.001$ e refaça o experimento.

Apresentar: Os valores de w

=> $W = [0.019534, 0.021172]$

Comentários: Comente sobre os valores de w obtidos

=> Os valores de W obtidos são baixos demais e a superfície de classificação não está ajustada o suficiente aos dados. Isso é causado pelo baixo valor de C, que ocasiona underfitting.

- Altere a classe do elemento 37 do conjunto de dados. Para isso, faça $y(37) = 1$. Treine o modelo utilizando $C = 1e9$

Apresentar: Figuras com o conjunto de dados e a superfície de separação

=> A execução 'crashou' o matlab após aproximadamente 10min.

Comentários: Comente sobre o tempo de convergência do método.

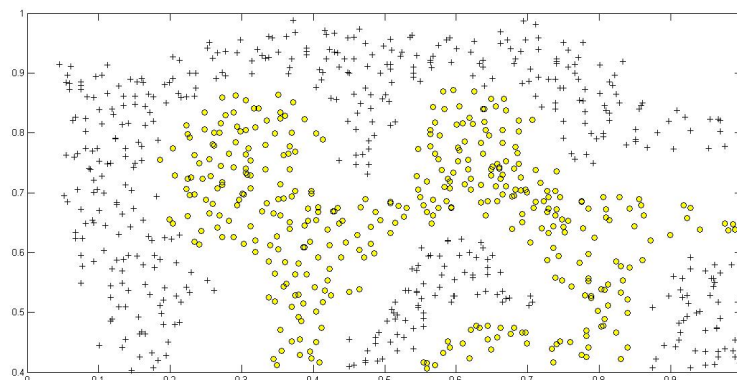
=> O método ficou executando por 10min e não convergiu. Isso foi causado porque o algoritmo só consegue calcular superfícies de classificação lineares e, nesse caso, é necessária uma não-linear.

2. Classificação usando SVM e Kernel RBF

- Carregue os dados contidos no arquivo ex6data2.data.

O arquivo contém uma matriz e um vetor de dados. A matriz X é composta de 863 linhas e 2 colunas e representa um conjunto de dados de dimensão 2. O vetor y dá a classe a qual pertence cada vetor. Este exemplo consiste em um problema de duas classes.

Apresentar: Figura com o conjunto de dados



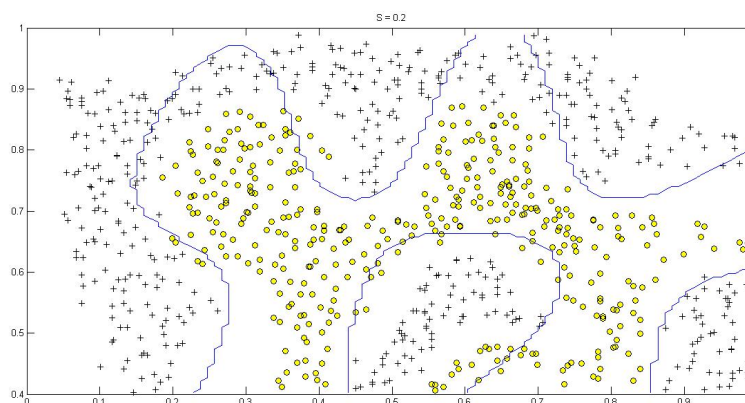
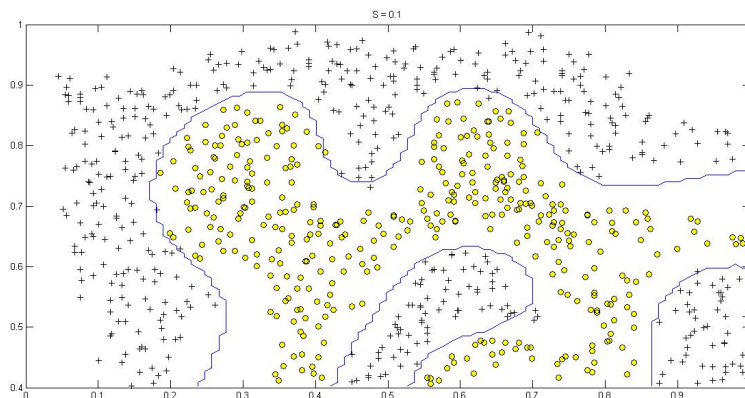
Comentários: Comente sobre qual tipo de Kernel deve ser utilizado neste problema.

=> Deve ser utilizado um Kernel não-linear nesse problema, pois não é possível encontrar um modelo linear que seja efetivo nesse conjunto de dados, uma vez que existem várias irregularidades na disposição das duas classes.

- Utilize a função `svmTrain` para treinar um modelo. Para esta função utilize o Kernel RBF. Utilize valores de $\sigma = 0.1$ e $\sigma = 0.2$. Utilize $C = 1$.

- Plote o resultado utilizando a função `visualizeBoundary`.

Apresentar: Figuras com o conjunto de dados e as superfícies de separação



Comentários: Comente sobre as superfícies de separação obtidas para os dois valores de sigma

=> No caso de $\sigma = 0.1$, o modelo obtido se adapta melhor ao conjunto de dados, uma vez que baixos valores de sigma diminuem a variância do modelo. Com $\sigma = 0.2$, obtemos um modelo mais irregular, com muitos outliers, que pode se adequar melhor ao adicionarmos mais exemplos.