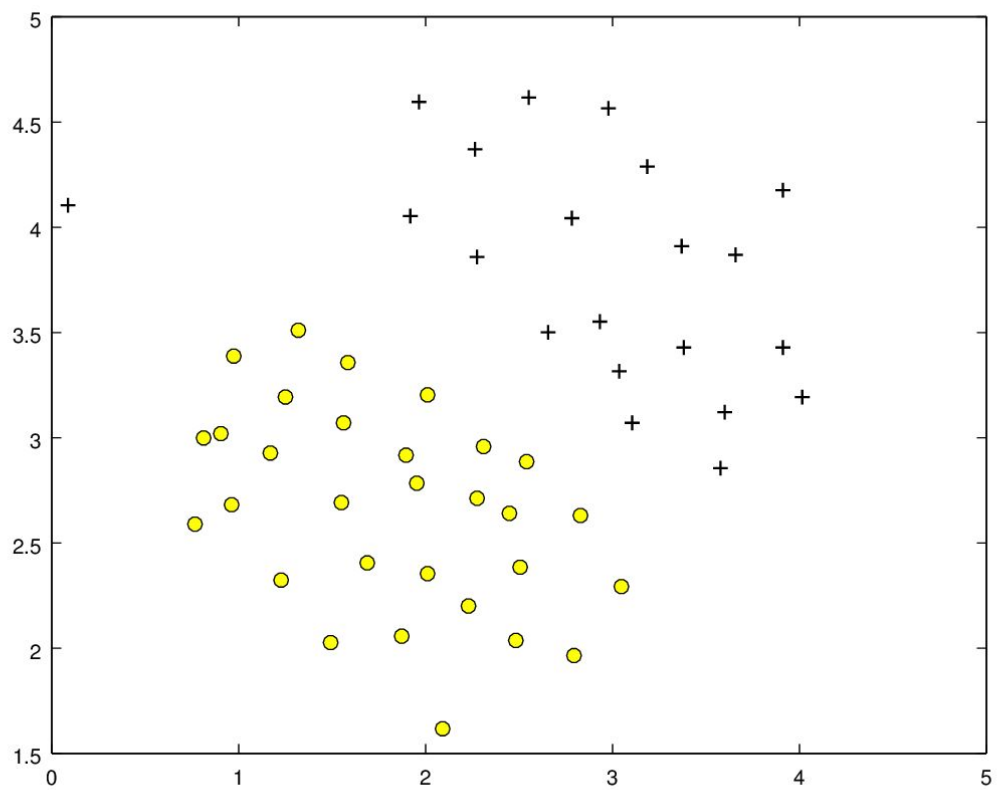


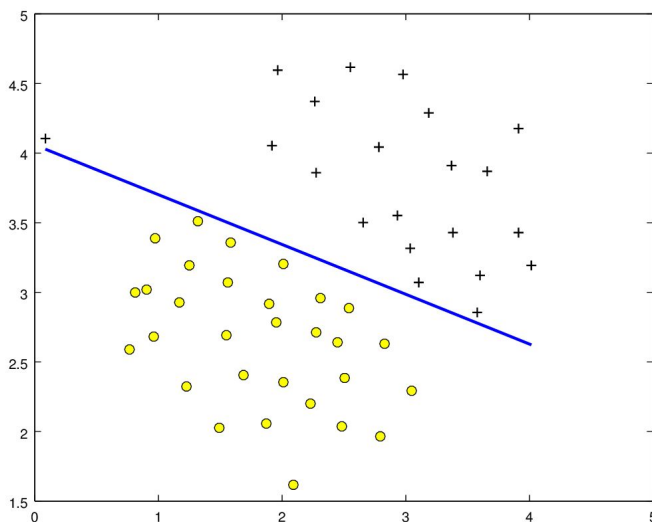
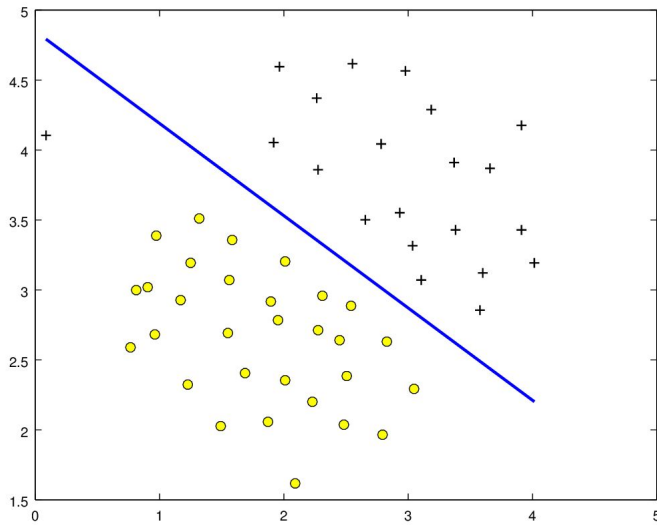
Paulo Bruno de Sousa Serafim - 388149
Aprendizagem Automática - Trabalho 5 - SVM

Questão 1.

Parte 1:



Parte 2:

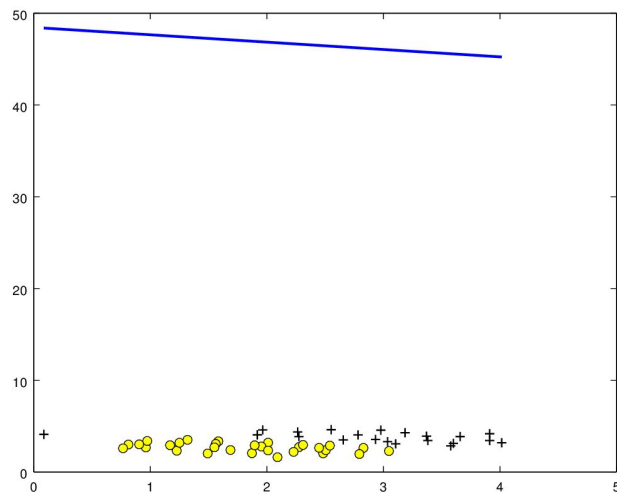


Comente sobre as superfícies de separação obtidas para os dois valores de C.

De maneira informal, C representa a “importância” da corretude da classificação, ou seja, quanto maior for C , mais se deseja que os dados sejam classificados corretamente. Inversamente, quando C for menor, admite-se uma certa intolerância na classificação em troca de uma superfície mais adequada.

Podemos visualizar que quando $C = 1$ a superfície de separação parece ser bastante adequada, entretanto há um valor classificado erroneamente. Já quando $C = 100$, espera-se uma maior precisão da classificação, mesmo que isso acarrete em uma superfície não muito adequada. Nesse caso, todos os dados foram classificados corretamente, entretanto a superfície não parece ser muito boa, ou seja, houve **overfitting**.

Parte 3:

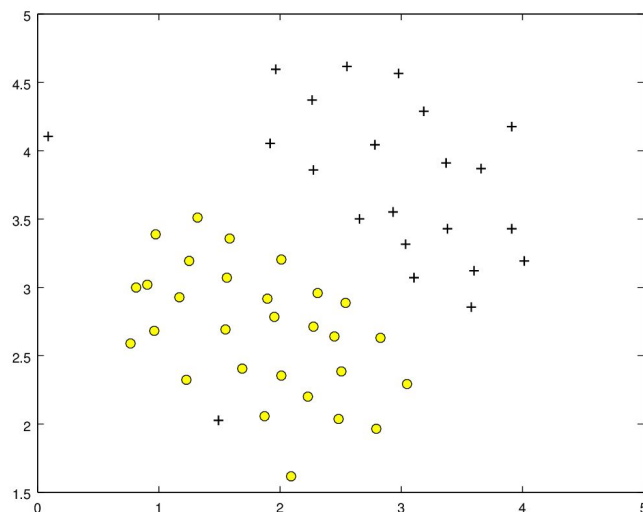


$$w = [0.018422 \ 0.021809]$$

Comente sobre os valores de w obtidos.

Como o valor de C é ínfimo, não há preocupação com a classificação correta dos dados. Dessa forma, o problema de otimização se resume a basicamente $\min \frac{1}{2} w^T w$. Assim, espera-se que os valores de w obtidos sejam de fato bem pequenos, devido à sua minimização. De fato, é o que ocorre.

Parte 4:

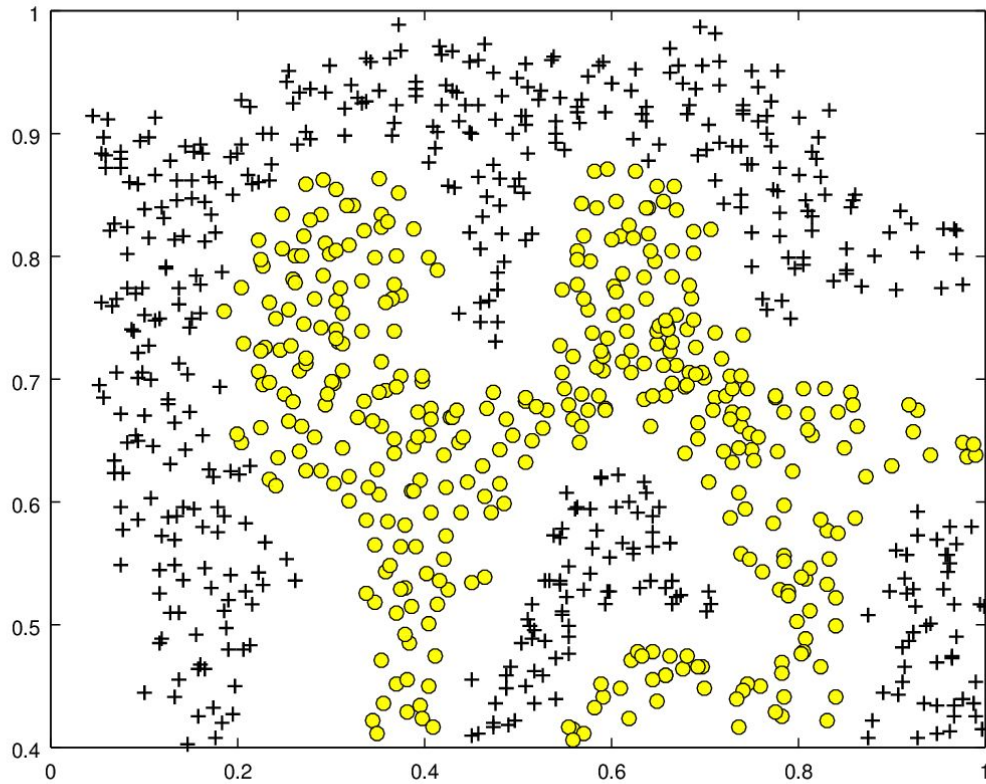


Comente sobre o tempo de convergência do método.

Como foi dito na Parte 2, para um valor muito alto de C, haverá uma tentativa de que todos os dados sejam classificados corretamente. Entretanto, como podemos observar na imagem acima, essa superfície não existe. Desse modo, o treinador tenta ajustar a superfície indefinidamente.

Questão 2.

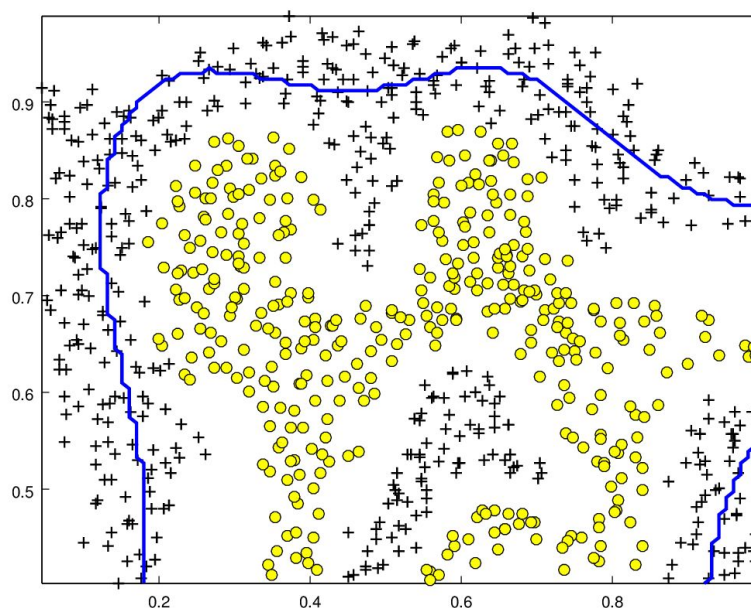
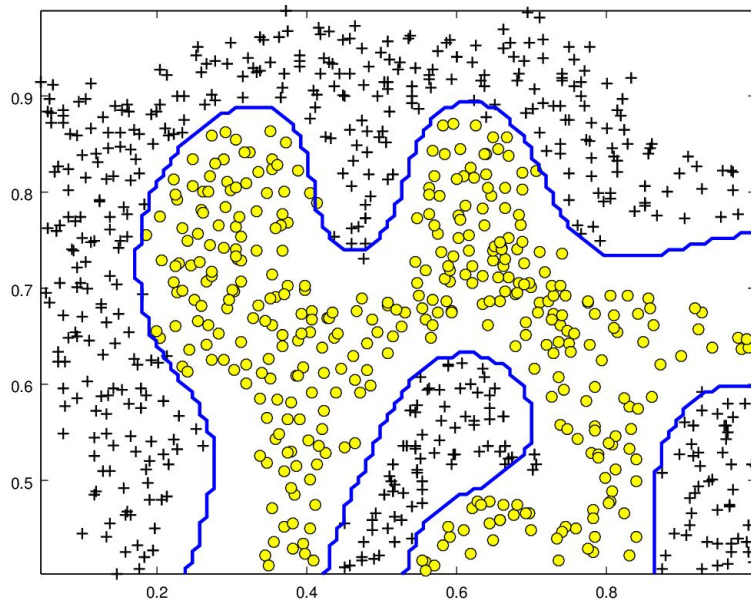
Parte 1:



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

Visualmente, podemos observar claramente que os dados não são linearmente separáveis, portanto será necessário utilizar um kernel não-linear, como o kernel RBF.

Parte 2:



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

Valores de sigma baixos, restringem a superfície aos dados, ou seja, dá mais importância à classificação dos dados. Já para um valor de sigma mais elevado, pode-se dizer de uma certa maneira que tenta capturar mais a forma da superfície de classificação, permitindo uma tolerância na classificação dos dados. Assim, podemos ver que enquanto no último caso a superfície parece bastante adequada, no primeiro podemos observar **overfitting**.