

Machine Learning for Computer Vision (EE5177)

Programming Assignment 4 : Logistic Regression and SVMs

Problem #2

Akshit Kumar
EE14B127

17th April 2017

Contents

1	Introduction	1
1.1	Goal	1
1.2	Approach	1
2	Linearly Separable Data	1
2.1	Bias-Variance Tradeoff	1
2.2	Contour Plots for different λ values	2
2.3	Required Iteration Data for $\lambda = 1$	4
2.4	Required Iteration Data for $\lambda = 10$	4
3	Non Linearly Separable Data : Use of Slack Variables	4
3.1	Goal	4
3.2	Approach	4
3.3	Required Iteration Data for $\lambda = 10$ and $C = 1$	4
3.4	Required Iteration Data for $\lambda = 10$ and $C = 100000$	4
3.5	Varying C	4
3.6	Plots for various values of C	4

1 Introduction

1.1 Goal

The goal of the second question is to use SVMs to find hyperplane between linearly separable datapoints. The optional exercise requires the use of slack variables to find an optimum separating hyperplane between non-linearly separable data.

1.2 Approach

For getting the separating hyperplane, we formulate the dual of the SVM optimization problem and pass it through the convex optimization library called cvx.

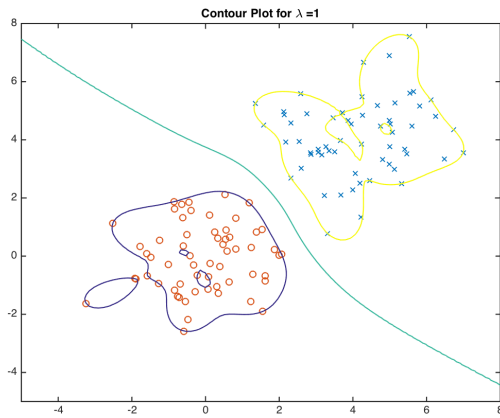
2 Linearly Separable Data

2.1 Bias-Variance Tradeoff

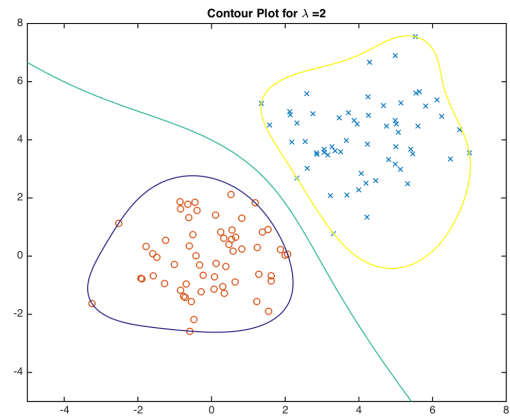
As we can clearly see from the plots for different values of λ , the lower the value of λ , the more the data is overfitted and number of parameters are higher. Thus the variance increases and bias reduces. This is because the model starts to become more and more complicated with more non-zero values of α . For $\lambda = 10$, we get 3 non-zero values of α which represents the number of support vectors used. For $\lambda = 1$, we get 32 non-zero values of α which represents the number of support vectors used. We see that higher the value of λ , the more general the model is and fits better to the test data compared to the models due to lower value of λ .

The optimal value of α vector obtained for $\lambda = 10$ has the following non-zero values : 102.769, 11.018 and 91.974

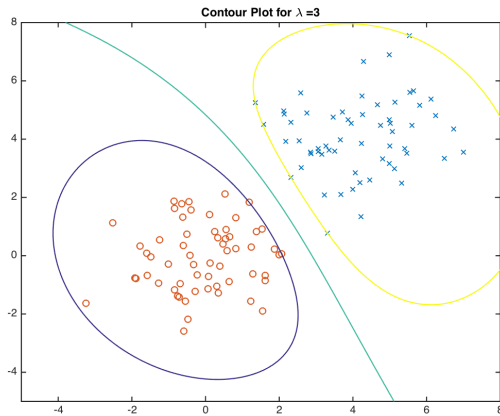
2.2 Contour Plots for different λ values



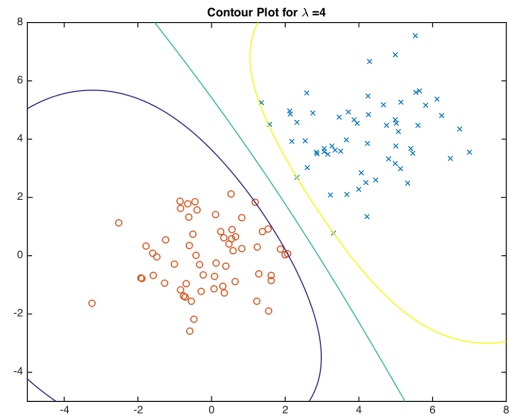
(a) Contour Plot for $\lambda = 1$



(b) Contour Plot for $\lambda = 2$

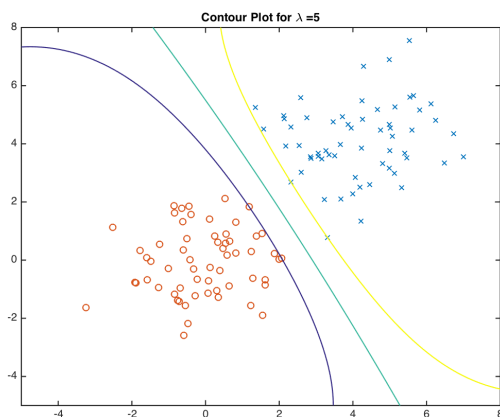


(c) Contour Plot for $\lambda = 3$

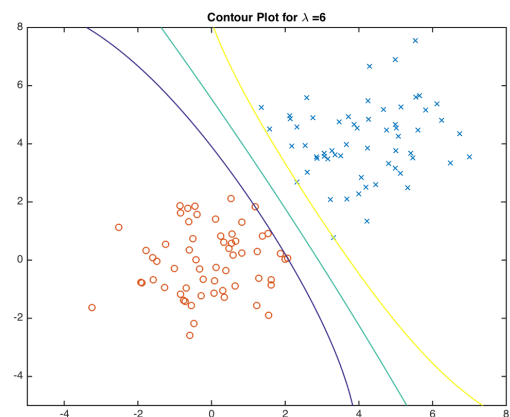


(d) Contour Plot for $\lambda = 4$

Figure 1: Small value of λ

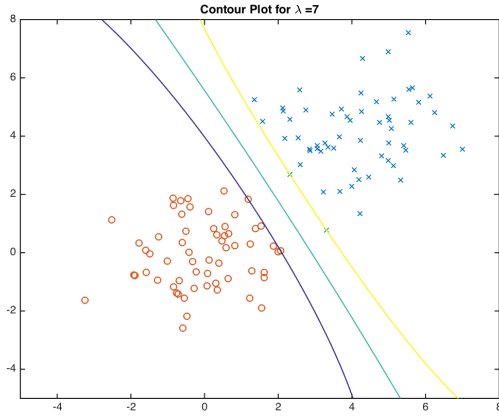


(a) Contour Plot for $\lambda = 5$

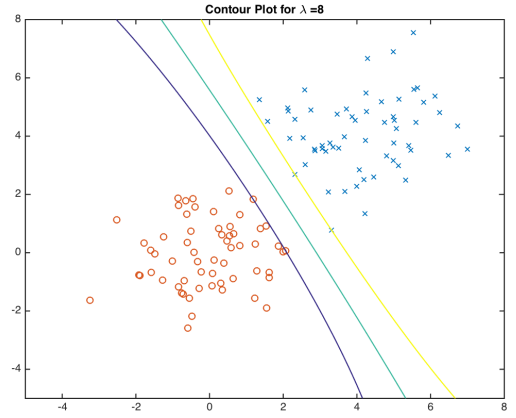


(b) Contour Plot for $\lambda = 6$

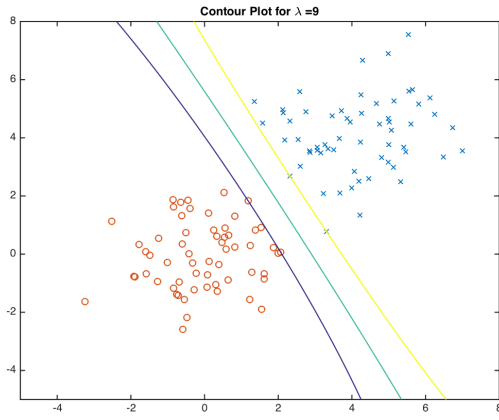
Figure 2: Moderate value of λ



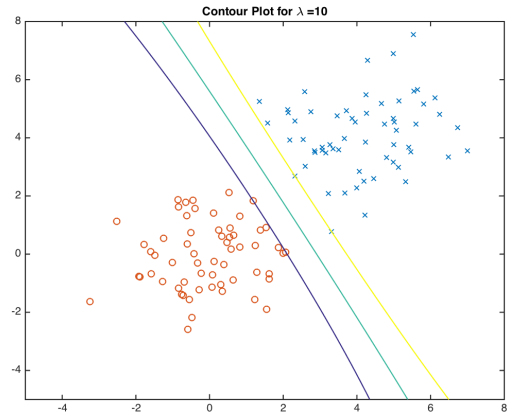
(a) Contour Plot for $\lambda = 7$



(b) Contour Plot for $\lambda = 8$



(c) Contour Plot for $\lambda = 9$



(d) Contour Plot for $\lambda = 10$

Figure 3: Large value of λ

2.3 Required Iteration Data for $\lambda = 1$

- Number of iterations : 23
- $\text{norm}(X), \text{norm}(y), \text{norm}(Z) = 1.2\text{e}+01, 9.1\text{e}+00, 1.2\text{e}+01$
- $\text{norm}(A), \text{norm}(b), \text{norm}(C) = 1.7\text{e}+01, 1.2\text{e}+01, 2.0\text{e}+00$
- Total CPU time(secs) : 0.68
- Optimal value for α :-7.83293

2.4 Required Iteration Data for $\lambda = 10$

- Number of iterations : 15
- $\text{norm}(X), \text{norm}(y), \text{norm}(Z) = 1.5\text{e}+02, 1.1\text{e}+02, 1.5\text{e}+02$
- $\text{norm}(A), \text{norm}(b), \text{norm}(C) = 9.7\text{e}+00, 2.0\text{e}+00, 1.2\text{e}+01$
- Total CPU time(secs) : 0.40
- Optimal value for α : -102.882

3 Non Linearly Separable Data : Use of Slack Variables

3.1 Goal

In this part of the problem, we modify the above used convex optimization problem by introducing slack variables to classify data which is non linearly separable.

3.2 Approach

Since the dataset is not linearly separable, we are required to misclassify certain points to get an optimal solution. Hence we add slack variables to generate the separating hyperplane.

3.3 Required Iteration Data for $\lambda = 10$ and $C = 1$

- Number of iterations : 18
- $\text{norm}(X), \text{norm}(y), \text{norm}(Z) = 2.0\text{e}+01, 7.6\text{e}+00, 1.2\text{e}+01$
- $\text{norm}(A), \text{norm}(b), \text{norm}(C) = 1.8\text{e}+01, 1.2\text{e}+01, 1.2\text{e}+01$
- Total CPU time(secs) : 0.55
- Optimal value for α : -28.1351

3.4 Required Iteration Data for $\lambda = 10$ and $C = 100000$

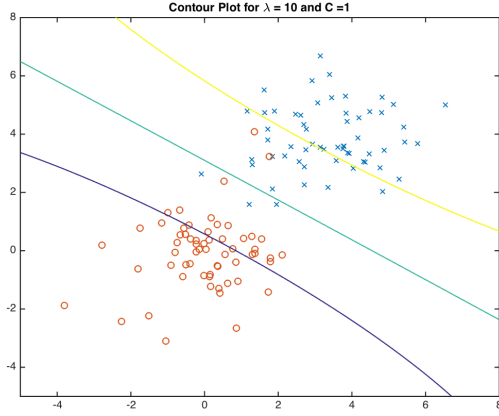
- Number of iterations : 25
- $\text{norm}(X), \text{norm}(y), \text{norm}(Z) = 3.5\text{e}+04, 2.6\text{e}+05, 1.1\text{e}+06$
- $\text{norm}(A), \text{norm}(b), \text{norm}(C) = 1.8\text{e}+01, 1.2\text{e}+01, 1.1\text{e}+06$
- Total CPU time(secs) : 0.70
- Optimal value for α : -674174

3.5 Varying C

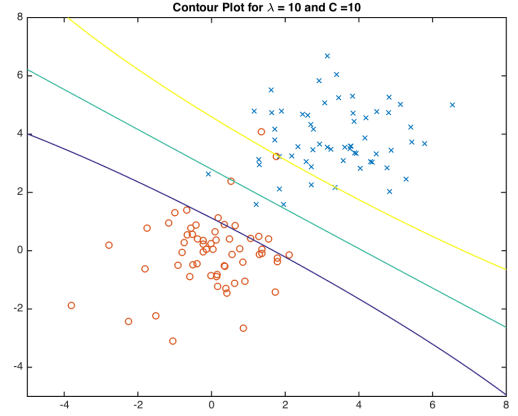
The parameter C controls how much we penalise the objective function for misclassifying a data point. For large values of C, the model would not be willing to misclassify any data point. For small values of C, the model would misclassify data very easily.

In case of a dataset which is not linearly separable, it is better to not overfit the data and use smaller value of slack. This should capture the general trend in the data and overfit the same.

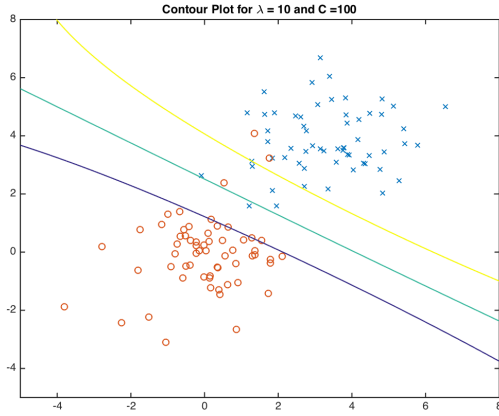
3.6 Plots for various values of C



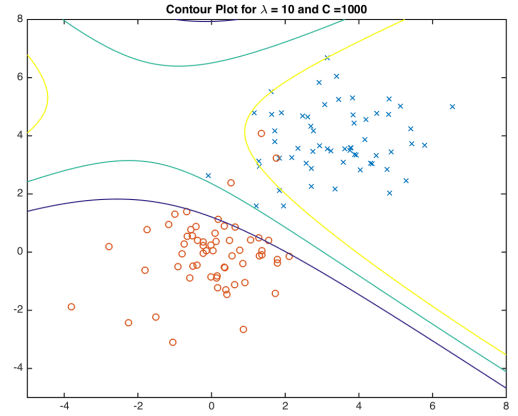
(a) Contour Plot for $\lambda = 10$ and $C = 1$



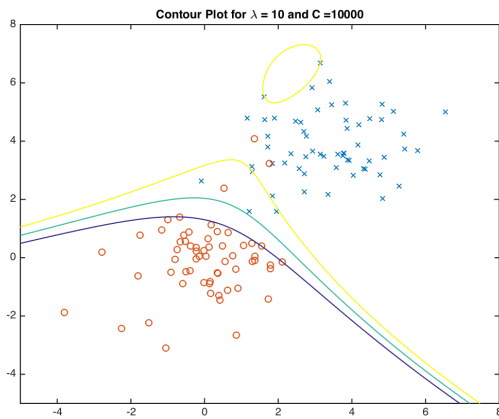
(b) Contour Plot for $\lambda = 10$ and $C = 10$



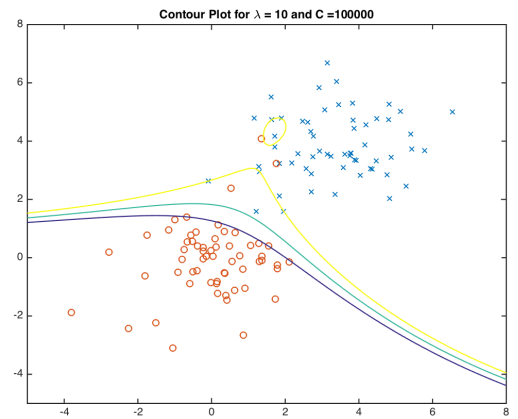
(c) Contour Plot for $\lambda = 10$ and $C = 100$



(d) Contour Plot for $\lambda = 10$ and $C = 1000$



(e) Contour Plot for $\lambda = 10$ and $C = 10000$



(f) Contour Plot for $\lambda = 10$ and $C = 100000$