Homework3.2

Susan Sapkota

December 2020

1 Result

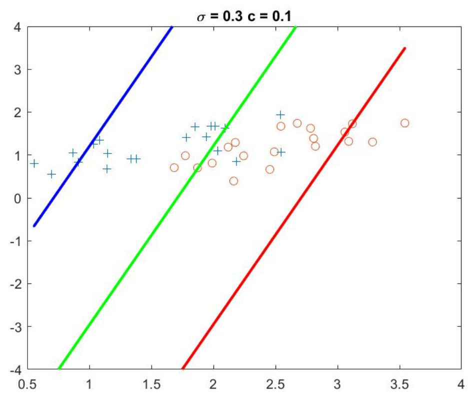

Figure 1: Classifier with $\sigma = 0.3$ and C=0.1

Figure 1 to 8 represents task 1 and 2 with varying parameter C and standard deviation σ. we can see that on figure 1,2,3 and 4 with fixed $\sigma = 0.3$ and increasing cost parameter (C). The classifier with high C have narrower margin and with low c has wider margin. The structural risk is dominant in case of low C whereas empirical risk is dominant in case of high C. So, Optimal classifier is on some medium value of C. Furthermore, we can see on 5,6,7 and 8 with fixed $\sigma = 1$ and varying C. We relative choose higher σ where relative increase in C doesnot narrow the margin like of the previous figure but also the margin gets narrower when increasing the parameter C. We see clearly the data are less spread for lower sigma in figure 1,2,3 and 4. However, Data are more spread for higher σ in figure 5,6,7 and 8. The support vector increase in number on increasing the σ which is evident on comparing on figure 1 to 4 with low σ vs figure 5 to 8 resulting on wrong classification of data. So, Error increases with increasing σ. Figure 9 show behaviour of training error and test error with cost parameter C. For the low value of C, the structural error is dominant than empirical error. On increasing C to 1 or 2, we can see that both empirical and structural error has equal contribution to error where the optimal value of parameter exists. On further increasing C, we see that the empirical error is dominant.

The complexity is low when the value of c is lower which leads to under fitting. On increasing C, the complexity is higher resulting test data poorly classified and over-fitting. We know risk can only be upper bounded so there is no physical or numerical way to calculate structural risk.

The structural risk is bounded by total risk and empirical risk by equation 3. Theoretical model predicts the actual risk is double of empirical risk where both meets but it doesnot intersect in practice. On increasing N on figure 10 (C=0.1), both test and training error goes to zero quickly. On increasing N on figure 11 (C=10), the test error decreases quickly and training increases because the dimension of the classifier is held fixed at 10. From the VC theorem, we can only shattered 11 points. On increasing the number of points, the classifier is not good enough to shattered point and training error increases. With increasing number of points, there is support vector with large margin which might be the reason for decreasing the test error.

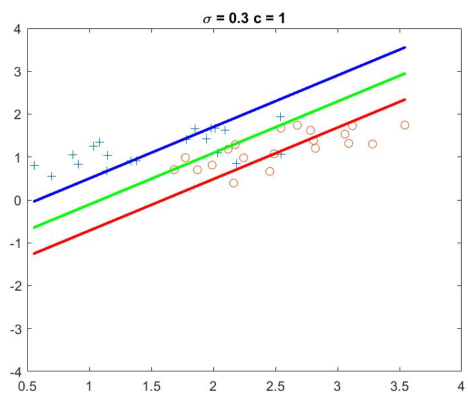

Figure 2: Classifier with $\sigma = 0.3$ and C=1

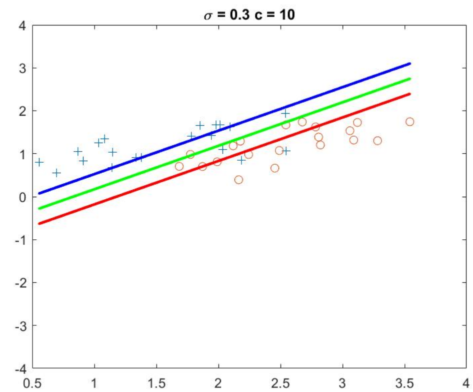

Figure 3: Classifier with $\sigma = 0.3$ and C=10

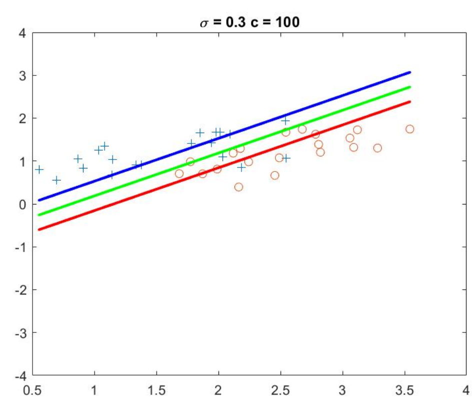

Figure 4: Classifier with $\sigma = 0.3$ and C=100

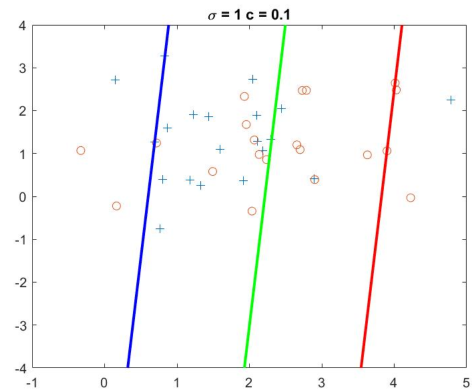

Figure 5: Classifier with $\sigma = 1$ and C=0.1

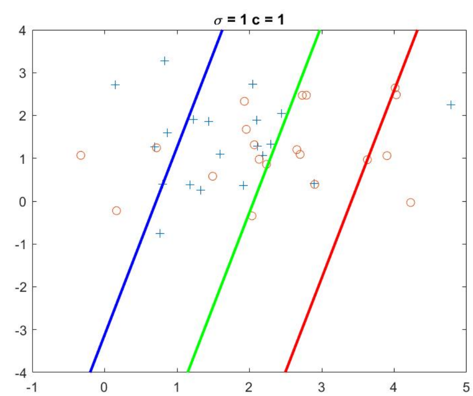

Figure 6: Classifier with $\sigma = 1$ and C=1

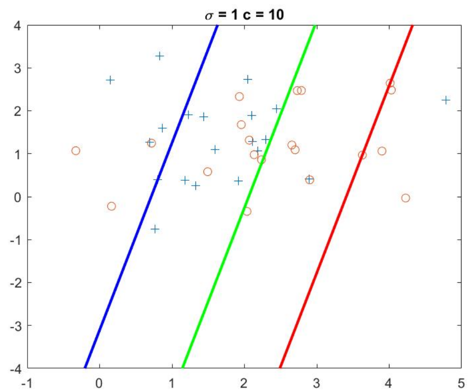

Figure 7: Classifier with $\sigma = 1$ and C=10

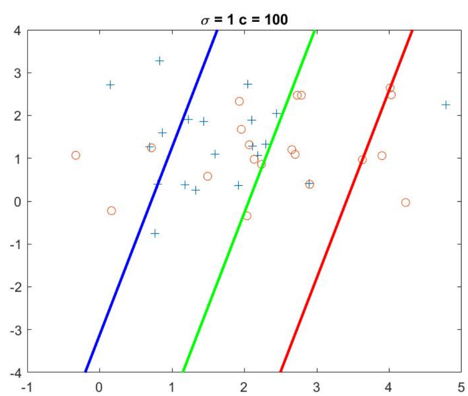

Figure 8: Classifier with $\sigma = 1$ and C=100

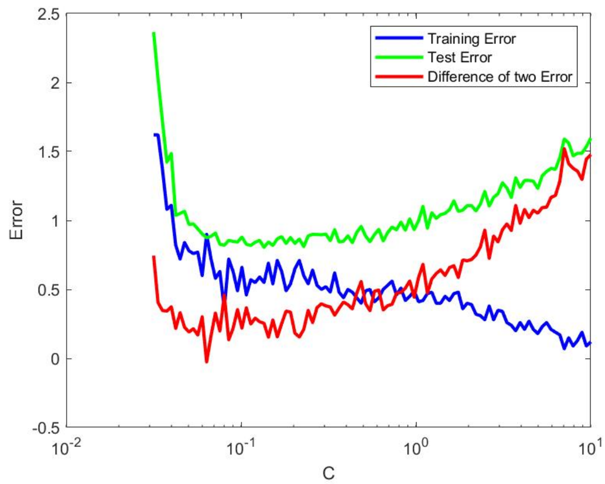

Figure 9: Variation of Error with Cost and N=100

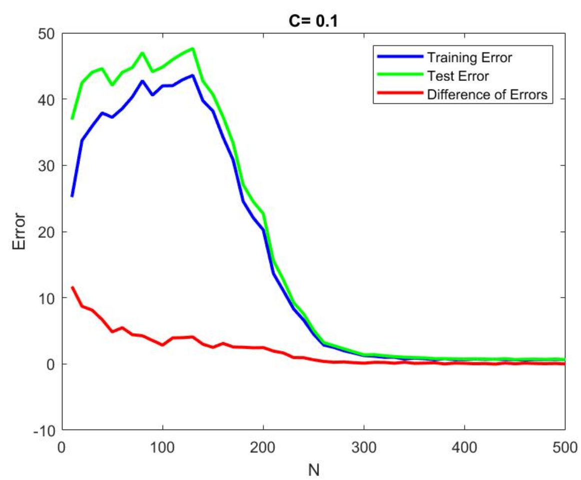

Figure 10: Variation of Error with N and fixed C=0.1

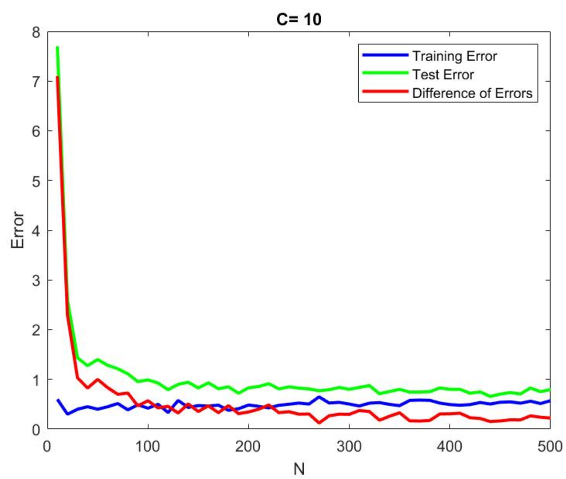

Figure 11: Variation of Error with N and fixed C=10