Support Vector Machine

Sai Krishna Kalyan

14th December 2015

1 Introduction

One of the most prevalent applications of convex optimisation methods in machine learning is the support vector machine classifier. Finding the support vector classifier can be formulated as the optimisation problem below.

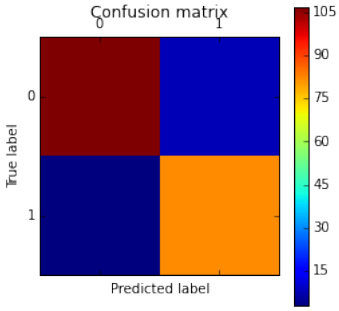
$$\min_{(w,\gamma,y)\in\mathbb{R}^{n+1+m}} \nu e^T y \ + \frac{1}{2} \ ||w||_2^2$$
$$s.t \quad D(Aw - e\gamma) + y \geq e$$
$$y \geq 0$$

2 Scope

To classify m points in $\mathbb{R}^n$, represented by matrix A $\in \mathbb{R}^{m\times n}$.Each point belongs to the class +1 or -1 depending on the classification of its diagonal matrix D $\in \mathbb{R}^{m\times m}$. For this assignment we will implement SVM with a linear kernel.

3 Convex Optimization

Numpy and cvxpy has been used to model this optimisation problem. Using gensvmdat.exe We have generated 2,00,000 data points. Subset of this data in $\mathbb{R}^{1000\times 5}$ has been used to train linear SVM classifier and another subset in $\mathbb{R}^{200\times 5}$ has been used to test the accuracy of the classifier.
Regularisation parameter ν takes in a scalar value greater than 0. Value of ν based on classification accuracy. (In this experiment we see that ν at value 0.6 gives us the greatest classification accuracy).
w,γ,y are the parameters that need to be optimised. D is a diagonal matrix of y and e is vector of ones.
To measure the performance of this experiment i have used metrics like precision, recall and F1 score. Their corresponding values observed were $0.97, 0.93, 0.95$. Around 189 points were correctly classified and 11 points were misclassified.



4 Conclusion

SVM performs much better over data sets with high dimensionality here even though we had only 4 dimensions, The algorithm classified data points quite well.
Tuning the regularisation parameter over the test data set prevents over fitting and gave us better generalised model.