Binary Classification based on Support Vector Machines

Prabhneet Arora

Abstract - Support Vector Machines also known as Support Vector Networks [1], are supervised learning models with associated learning algorithms that help in analyzing data used for classification and regression analysis. This technique has been employed to find the best kernel for Binary classification of the data at hand. Four different kernels have been compared to get the maximum classification accuracy. Hyper-parameter tuning has been performed to achieve the best values for C and gamma corresponding to greatest classification accuracy, which are our tuning parameters.

Index Terms - Support Vector Machines, Binary Classification, kernel, hyperparameters

I. INTRODUCTION

Dataset consisting of two features i.e. Feature 0 and Feature 1, generated by scikit learn feature make_classification was used for binary classification. The initial dataset is represented by Fig. 1.

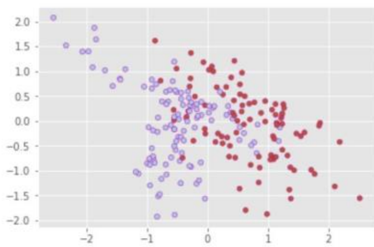

Fig. 1 Graphical representation of initial dataset with the two classes depicted in two different colours.

II. METHODOLOGY

The decision surface articulated from four SVM Classifiers with different kernels was used for our analysis. They are linear SVC, SVC with linear kernel, SVC with RBF kernel and 3rd degree polynomial kernel. Both linear models have linear decision boundaries, while non-linear have more flexible decision boundaries with shapes that greatly depend on the type of kernel used and its myriad parameters. For the RBF kernel, C and gamma are the tuning parameters, gamma defines the radius of influence of the support vectors, whereas C trades-off the correct classification of training examples against the maximum margin of the hyperplane. Grid search was further employed to find the best combination of C and gamma for the RBF kernel that yielded the highest score.

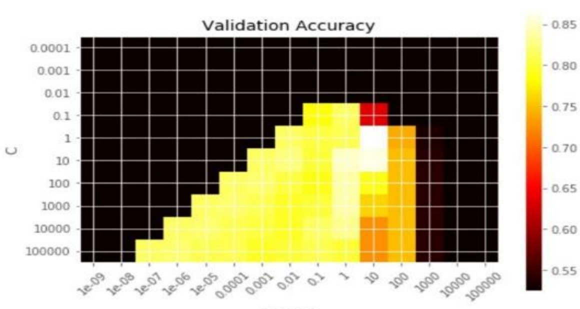

Fig. 2 Heat Map of validation accuracy (gamma vs C values). The highest accuracy is colored the lightest as evident from the scale on the right.

II. RESULTS AND DISCUSSION

SVM model behavior with RBF kernel is very sensitive to the gamma and C values. Since gamma captures the complexity of the model, choosing a good gamma value proved to be an essential component to avoid overfitting of training examples. Using Grid search we found out that values of C and gamma at 1 and 10 respectively, yielded the highest score of 0.87 as can be seen in Fig. 2.

Although, linear kernels tend to be much faster, but the predictive performance of the non-linear kernels especially RBF was found out to be better than the linear kernels as shown in TABLE I. Performance comparison between the four kernels can be seen in Fig. 3. All three kernels resulted in lower classification score that RBF as can be seen in Table 1.

It has been revealed in research studies that linear kernel is a degenerate version of RBF or Gaussian kernel [2], hence properly tuned RBF kernel tends to perform well and yields more accurate results than linear kernel.

In practice, linear kernels perform better when the number of features is large (i.e. in a high dimensional feature space). Non-linear kernels are not significantly more accurate than linear so they lose their appeal since they require unnecessary tuning of many parameters with little gain in performance.

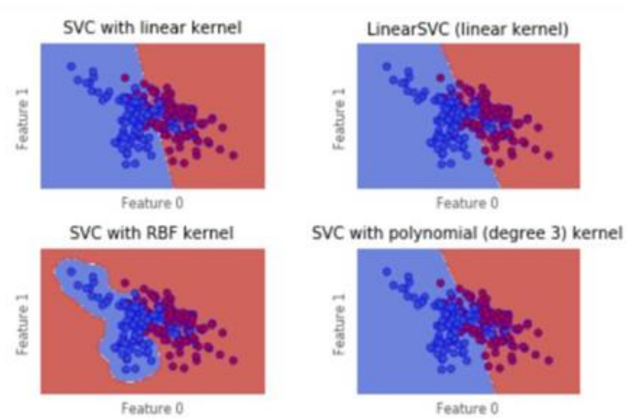

Fig. 3 Binary classification classified according to the four kernels.

TABLE I

S. No.	Kernel Type	Score
1	Linear	0.83
2	Linear SVC (linear kernel)	0.81
3	RBF (C=1, gamma =10)	0.87
4	Polynomial (degree 3)	0.83

REFERENCES

[1] Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine learning, 20(3), 273-297.

[2] Keerthi, S. S., & Lin, C. J. (2003). Asymptotic behaviors of support vector machines with Gaussian kernel. Neural computation, 15(7), 1667-1689.