ECEN 689
Challenge 5: Support Vector Machines

Mason Rumuly

Abstract

The focus of this project is to choose the kernel which classifies the provided dataset with the best generalizability. Of the kernels considered, keep-one-out cross-validation showed the linear kernel performs best on the provided set.

I. Introduction

SUPPORT Vector Machines (SVM) became popular because of efficient computation using quadratic programming allowing them to outperform practical artificial neural networks (ANN) when speed is a concern. The kernel trick allows it to apply directly to datasets with certain non-linear optimal decision boundaries, restricted by the computational tractability of the kernel. This challenge focuses on the choice of kernel for optimal performance of an SVM classifier.

II. Methods

The Scikit-Learn Support Vector Machine Classifier (SVC) served as the SVM implementation for this task [1]. Due to the manageable data set size, keep-one-out cross validation was used to assess each kernel's generalization ability on the dataset.

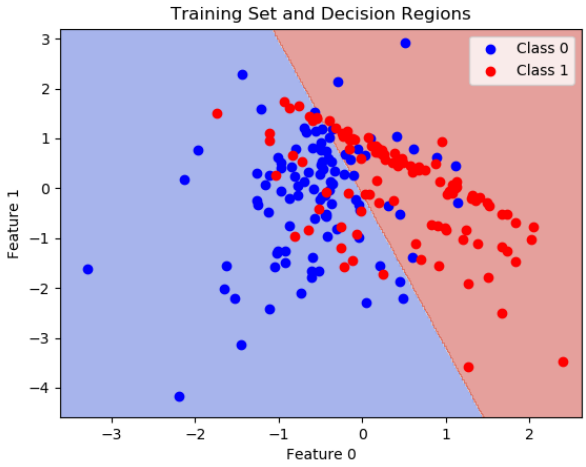
The kernels chosen for evaluation were the linear kernel, polynomial kernels of degree between 2 and 5, the radial basis function (RBF) kernel, and the sigmoid kernel.

III. Results

The Linear kernel performed best in cross validation, as seen in Table 1a. The decision regions resulting from training over the entire training set is shown in Figure 1b. This corresponds intuitively to the visible distribution of the training set. Note that the accuracy never breached 79% in cross-validation; this too fits with the intermixing of the classes visible in the plot.

Kernel	Cross Validation Accuracy
Linear	0.79
Quadratic	0.61
Cubic	0.695
Polynomial $d = 4$	0.595
Polynomial $d = 5$	0.615
RBF $d = 5$	0.765
Sigmoid	0.775

(a)



(b)

Fig. 1. (a): Keep-one-out cross validation results for each kernel (b): Feature-space plot of training samples and the decision region of the optimal kernel trained over the whole training set

IV. Conclusion

The kernel choice was completed successfully, yeilding good experience with the use of kernels with a support vector machine classifier.

References

[1] Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011.