ECEN 689 - Applied Information Science - Challenge 5

Harish Kumar

Abstract

This short report describes my experiments with SVMs for Challenge 5. I used four different kernel functions and performed a cross-validated grid search over a wide range of hyperparameters and identified the optimal kernel and its parameters.

I. Problem Description

THE problem is a simple binary classification task, with the training data containing 200 datapoints. We are expected to train and use a Support Vector Machine to classify 200 other test points. We first note that the dataset is not linearly separable, and hence expect the linear kernel to be somewhat poor in performance as compared to more complex kernels.

II. Kernel Performance

The following table describes the performance of various kernels on the dataset. The respective hyperparameters for all kernels were finely tuned with grid-search and 5-fold cross-validation. I used the GridSearchCV API in Scikit-learn to this purpose. Only the best performance for each kernel is shown in the below table.

Kernel	Best Performance	Hyperparameter Values
Linear	0.785	$C = 10.0$
RBF	0.83	$C = 8.002, \gamma = 0.1$
Sigmoid	0.79	$C = 25.11, \gamma = 0.0158$
Polynomial	0.745	$C = 398.10, degree = 3, \gamma = 0.1$

Based on this, the best performance is offered by the RBF Kernel with parameters $C = 8.002, \gamma = 0.1$. The cross-validation accuracy reported for these parameters is 0.83.

III. Decision Boundary

The decision boundary for the optimal SVM Classifier is shown in **Fig 1**. We do see that the boundary accurately envelops most of the purple points and carefully leaves out the yellow points. However, it is unable to capture the central cluster of purple points. During experiment, I noticed that polynomial kernels could sometimes capture this cluster, but this came at great cost to accuracy.

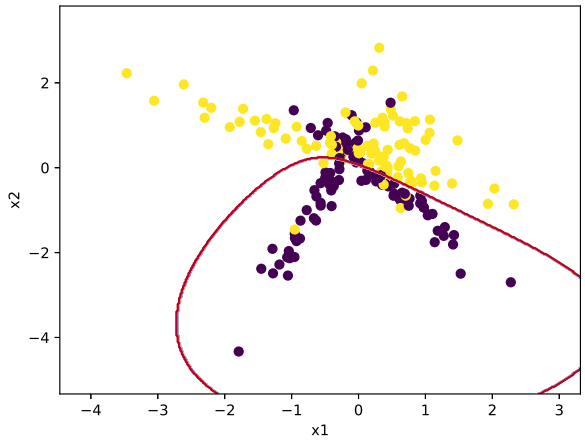

Fig. 1. The Decision Boundary for the optimal SVM Classifier