

Challenge 5: Support Vector Classifier

AliAkbar Shafi

Introduction

Challenge 5 is a binary classification problem. The challenge demanded the design of a support vector classifier to identify the class (0/1) of an observation with two features. Different classifiers were tried and tested on training set to determine the best fit kernel on the training data. Also, the decision boundary has been plotted based on the choice of kernel.

Procedure

The training data with 2 features is depicted in the figure 1. From the figure we cannot determine the exact boundary of classification, hence using a Support vector classifier with a flexible spline is an acceptable choice for determining the binary class of data.

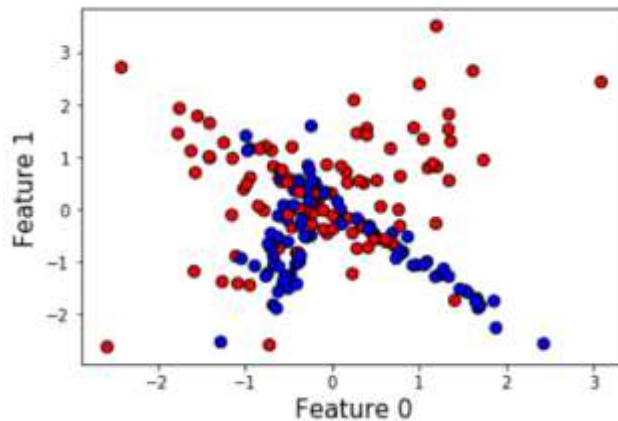


Figure 1

The SVC from scikit was used to implement the support vector machine. Scikit has 4 different kernels: linear, polynomial, rbf and sigmoid. Also, there are other parameters like degree of polynomial kernel or regularization of rbf kernel (c value) which can be hyper tuned to achieve a better fitting kernel.

To identify the best kernel and its parameters, grid search was applied using a GridSearchCV function where the classifier was run on training data with four kernels and multiple values of polynomial degree, regularization values. The best kernel for my data came out to be rbf with regularization value of 5. Also before applying the SVC on Training data, the data was split into training and validation data set. Figure 2 below shows the decision boundary of classification on complete training dataset.

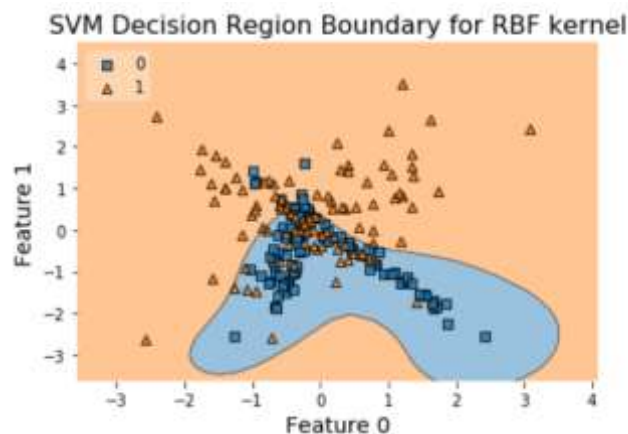


Figure 2