

Report: Challenge 5

Samyuktha Sankaran

November 8, 2018

Abstract

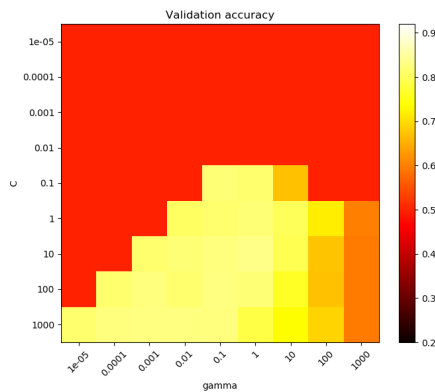
To build a binary classifier based on Support vector machine for the give data.

1 Model and Results

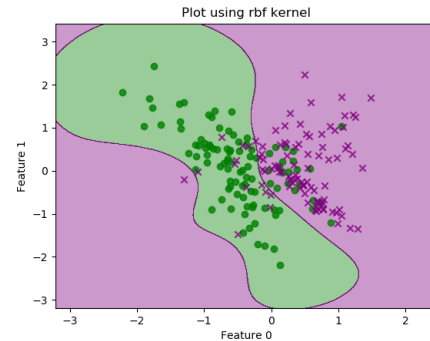
The given dataset comprises of two features and two classes. There are four different types of model available on SciKit: Linear, Polynomial, Radial Basis Function(RBF) and Sigmoid functions. The training data is split into two parts, the training and the validation set. On applying the four models to the training set and testing on the validation set, the RBF kernel showed the best clas-sification results with an accuracy of 85%. The following table shows the accuracy of different models

Kernel	Linear	Polynomial	RBF	Sigmoid
Accuracy	77.65%	70%	87.5%	67.5%

Heat map [Fig.1(a)], plotted between features for the dataset, show the co-relation between the C and gamma value. GridSearchCV function was used to find the best fitting parameter for the rbf kernel. The values for 'C' and γ ranges from 10^{-5} to 10^3 . The best parameters estimated by the grid search has $C = 10$ and $\gamma = 1$ with a score of 0.83. On trying to estimate the best value for these hyper-parameters using brute force resulted in $C = 1$ and $\gamma = 0.1$, but the accuracy dropped to 77.85%. Figure 1(b) shows the SVM decision regions for the data points separated into two classes $\{0,1\}$.



(a) Heat Map



(b) Training data

Figure 1