Project: #3
Jaswinder kaur

a. Classification results and accuracy:

Training	set	Accuracy:81.364%
Validation	set	Accuracy:80.28999999999999%
Testing	set	Accuracy:80.69%

Here I have approximately 80% accuracy which means the data is not linearly seperable and ha ve high input features.

b. Support Vector Machine:

A Support Vector Machine is a discriminative classifier formally defined by a separating hyperplane. In other words, given labeled training data the algorithm outputs an optimal hyper-plane which categorizes new examples. If the input features are high it works good.

1. Using linear kernel (all other parameters are kept default):

   training accuracy = 97.286%
   validation accuracy = 93.64%
   testing accuracy = 93.78%

Here I see using linear kernel the training accuracy lower than using default gamma.

2. Using radial basis function with value of gamma setting to 0.1

   training accuracy = 99.992%
   validation accuracy = 94.76%
   testing accuracy = 94.96%

Here the testing accuracy is greaterthan linear kernel and also slightly better than radial basis d efault gamma. And the training is higher too.

3. Using radial basis function with value of gamma setting to default (all other parameters are kept default)

   training accuracy = 94.294%
   validation accuracy = 94.02000000000001%
   testing accuracy = 94.42%

Here I see validation accuracy and testing accuracy is better than linear kernel and gamma0.1

4. Using radial basis function with value of gamma setting to default and varying value of C (1, 10, 20, 30, · · ·, 100) and plot the graph of accuracy with respect to values of C in the report:

Gamma, C = 1:

training accuracy = 94.3% validation accuracy = 94.02%testing accuracy = 94.42%

Gamma, C = 10:
training accuracy = 97.13199999999999% validation accuracy = 96.17999999999999%
testing accuracy = 96.1%

Gamma, C = 20:
training accuracy = 97.952% validation accuracy = 96.899%
testing accuracy = 96.67%

Gamma, C = 30: training accuracy = 98.37%validation accuracy = 97.1%
testing accuracy = 97.04%

Gamma, C=40:
training accuracy = 98.71%validation accuracy = 97.23%
testing accuracy = 97.19%

Gamma, C=50:
training accuracy = 99.002% validation accuracy = 97.31%
testing accuracy = 97.19%

Gamma, C=60:
training accuracy = 99.196% validation accuracy = 97.4%
testing accuracy = 97.16%

Gamma, C=70:
training accuracy = 99.34%validation accuracy = 97.36%
testing accuracy = 97.26%

Gamma, C = 80:
training accuracy = 99.438% validation accuracy = 97.39%
testing accuracy = 97.33%

Gamma, C = 90: training accuracy = 99.542% validation accuracy = 97.36%
testing accuracy = 97.34%

Gamma, C = 100: training accuracy = 99.612% validation accuracy = 97.41%
testing accuracy = 97.4%

Here We are seeing all the accuracies are increasing we increase C. If we put lower C values larg e then large margin hyper-plane will be created. And large C value gives will increase weight an d will give us smaller margin hyper-plane. But also having too high values can cause over-fitting. And too small value can cause under fitting.

I should use linear kernel when number of features is larger than number of observations. W e should use Gaussian kernel when number of observations is larger than number of features. U sing SVM with Gaussian kernel gives higher accuracy.