Support Vector Machines

Using the Support Vector Machine to perform classification on MNIST data set. The SVM model was learned and computed the accuracy of prediction with respect to training, validation and testing datasets using the following parameters:

1. Linear kernel
2. Radial basis function with value of gamma setting to 1 Using radial basis function with value of gamma setting to default
3. Radial basis function with value of gamma setting to default and varying value of C (1, 10, 20, 30 ,, 100)

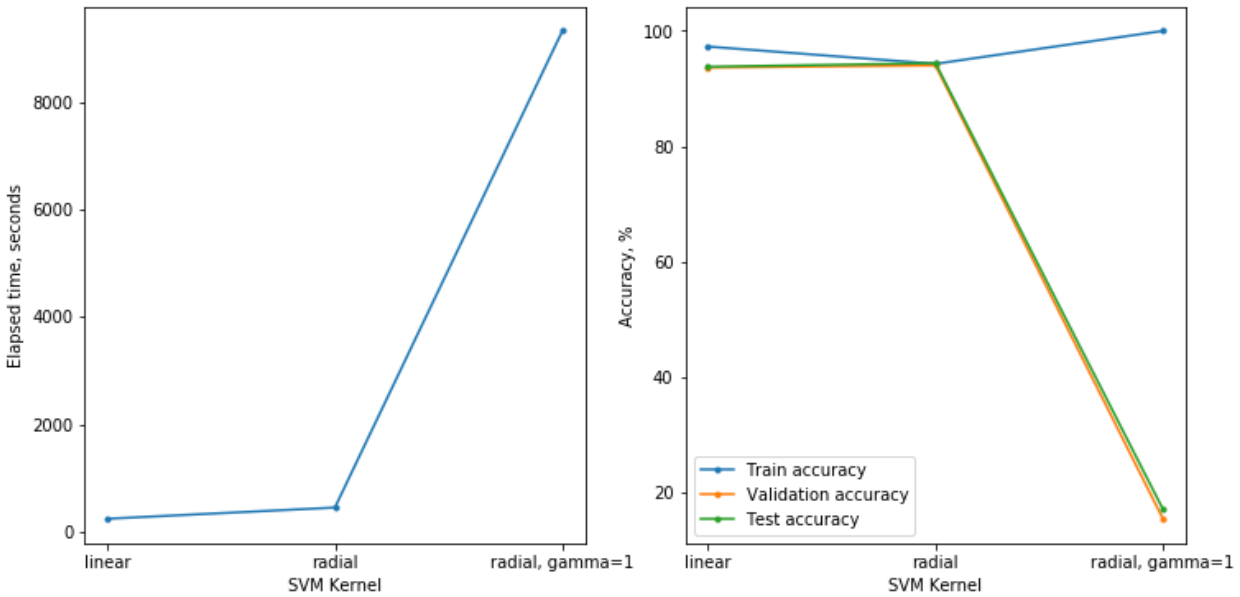
As a result, a graph of accuracy with respect to values of C was plotted.

Tools & Liblaries: Python, Numpy, Scipy, Sklearn

Report

SVMs with linear and radial basis kernels and a range of C values were trained on Metallica server against MNIST dataset.

Comparing SVM with linear kernel, radial kernels and radial kernel with gamma=1



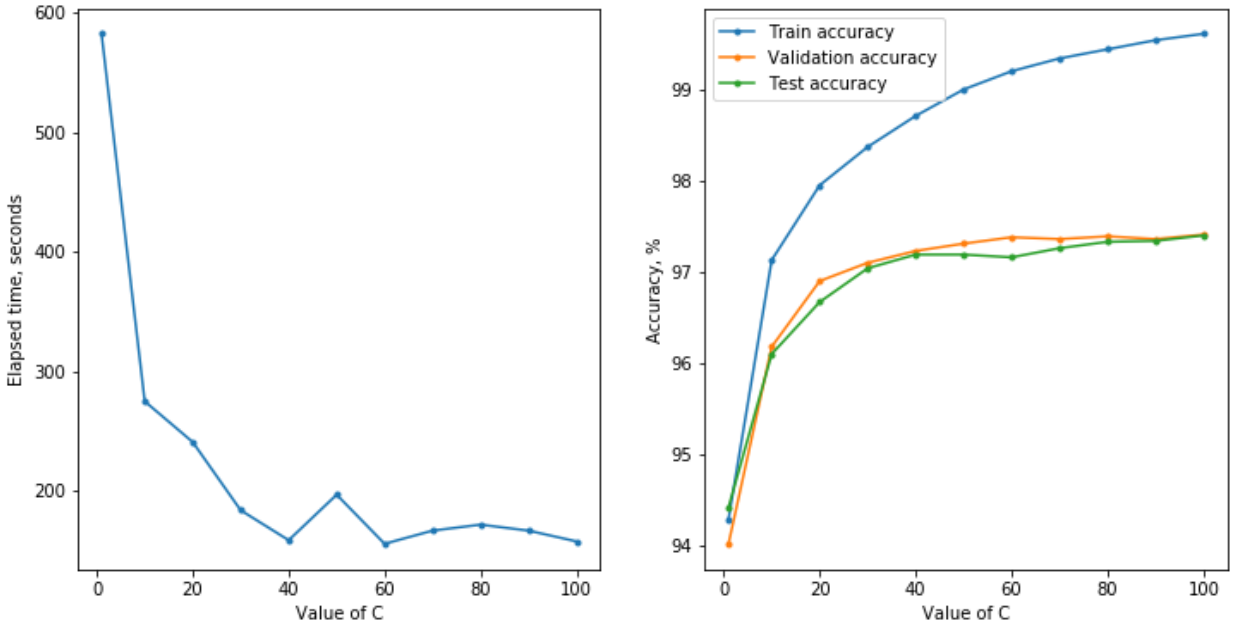
Linear kernel linearly separates the training samples, while radial kernel maps training dataset in a higher dimensional space, in which samples become linearly separable. Thus

radial kernel works good with non-linearly separable data, but it increases the model complexity, requires more computational time and we also have to adjust additional hyperparameters. As we can see from the results, the training time for linear model is 80% less than for radial kernel (253 seconds comparing to 461 seconds).

At the same time, the best result on testing dataset is given while using radial kernel, the accuracy result is 94.42%, comparing to 93.78% for linear kernel and 17.14% when $\gamma = 1$. Thus, if the accuracy is not a high priority we would recommend to use SVM with linear kernel, which is optimal in terms of accuracy for testing data and training time. This also shows that the MNIST dataset can be linearly separated with a relatively low error rate.

Also it took more that 2.5 hours to train SVM with radial kernel and $\gamma = 1$. It shows perfect result on the training dataset - 100%, but the lowest results on testing dataset (17.14%), which shows that there was a huge overfitting. γ regularizes the influence of points which are located further from the decision boundary. Thus when value of γ is high, it considers only the closest samples and assigns a high weights that will result in a perfect separation of the training set, with wiggly decision boundary, but it might result in a low accuracy for the testing data, as the model can be overfitted.

Comparing SVMs with radial kernel and different C



Parameter C controls the tradesoff between large decision boundary and classifying training points correctly. Thus when the value C is high - the decision boundary will be very tight, but it might not generalize well for unseen data. As we can see from the testing results, when $C = 100$ the model returns the highest accuracy for the training dataset (96.51%), along with the highest accuracy on testing dataset (94.7%). Also the training time when

$C = 100$ is one of the best among other C variations - 158 seconds.

On the contrary, when C value is low, it cares exclusively about maximizing the margin. It does not consider close single points, but more cares about the clusters of points with a high density. From our testing results, we can see that when $C = 1$ or $C = 10$ the model results with lower accuracy for all the datasets along with the highest training time.

Thus among C values for radial kernel we would recommend to use a relatively large C value to achieve better accuracy and lower training time.

This result is also the most optimal comparing to SVM with linear kernel, thus for this type of dataset it is better to use SVM with radial kernel and $C = 100$.

Accuracy results for SVM with different settings
————————-SVM with linear kernel————————
Time used: 253 seconds
Accuracy on Training dataset: 97.286%
Accuracy on Validation dataset: 93.64%
Accuracy on Testing dataset: 93.78%

————SVM with radial kernel, gamma=1————
Time used: 9317 seconds
Accuracy on Training dataset: 100.0%
Accuracy on Validation dataset: 15.48%
Accuracy on Testing dataset: 17.14%

————-SVM with radial kernel, gamma=auto————
Time used: 461 seconds
Accuracy on Training dataset: 94.29%
Accuracy on Validation dataset: 94.02%
Accuracy on Testing dataset: 94.42%

————-SVM with radial kernel, C=1————
Time used: 582 seconds
Accuracy on Training dataset: 94.29%
Accuracy on Validation dataset: 94.02%
Accuracy on Testing dataset: 94.42%

————-SVM with radial kernel, C=10————
‘Time used: 275 seconds
Accuracy on Training dataset: 97.13%
Accuracy on Validation dataset: 96.18%
Accuracy on Testing dataset: 96.1%

————-SVM with radial kernel, C=20————

Time used: 241 seconds
Accuracy on Training dataset: 97.95%
Accuracy on Validation dataset: 96.9%
Accuracy on Testing dataset: 96.67%

———-SVM with radial kernel, C=30———
Time used: 184 seconds
Accuracy on Training dataset: 98.37%
Accuracy on Validation dataset: 97.1%
Accuracy on Testing dataset: 97.04%

———-SVM with radial kernel, C=40———
Time used: 159 seconds
Accuracy on Training dataset: 98.71%
Accuracy on Validation dataset: 97.23%
Accuracy on Testing dataset: 97.19%

———-SVM with radial kernel, C=50———
Time used: 197 seconds
Accuracy on Training dataset: 99.0%
Accuracy on Validation dataset: 97.31%
Accuracy on Testing dataset: 97.19%

———-SVM with radial kernel, C=60———
Time used: 156 seconds
Accuracy on Training dataset: 99.2%
Accuracy on Validation dataset: 97.38%
Accuracy on Testing dataset: 97.16%

———-SVM with radial kernel, C=70———
Time used: 167 seconds
Accuracy on Training dataset: 99.34%
Accuracy on Validation dataset: 97.36%
Accuracy on Testing dataset: 97.26%

———-SVM with radial kernel, C=80———
Time used: 172 seconds
Accuracy on Training dataset: 99.44%
Accuracy on Validation dataset: 97.39%
Accuracy on Testing dataset: 97.33%

———-SVM with radial kernel, C=90———
Time used: 167 seconds
Accuracy on Training dataset: 99.54%
Accuracy on Validation dataset: 97.36%

Accuracy on Testing dataset: 97.34%

———-SVM with radial kernel, C=100———
Time used: 158 seconds
Accuracy on Training dataset: 99.61%
Accuracy on Validation dataset: 97.41%
Accuracy on Testing dataset: 97.4%