Homework 9 – SVM and AdaBoost

Name: Anoosha Sagar

Support Vector Machines

In machine learning, **support vector machines** are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier.

In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces.

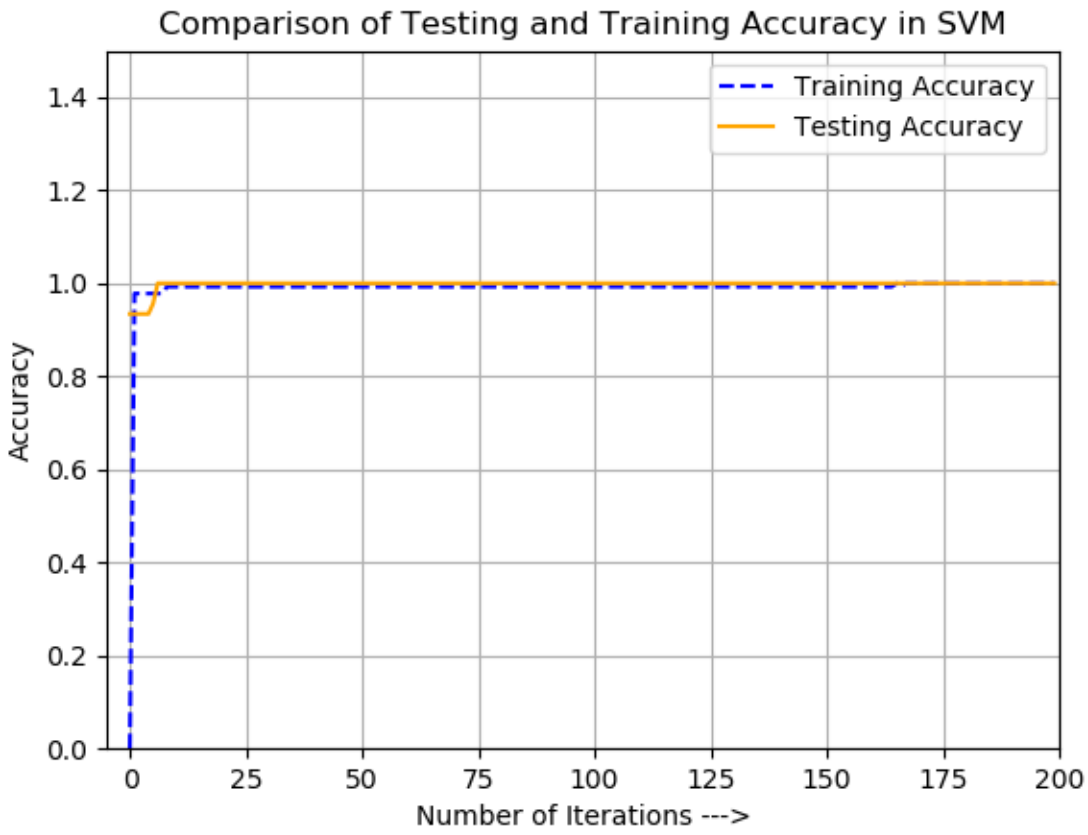
Plotting Training and Testing Accuracies

```
x, y, z = my_SVM(X_train, Y_train, X_test, Y_test)
plt.xlabel("Number of Iterations --->")
plt.ylabel("Accuracy")
plt.title("Comparison of Testing and Training Accuracy in SVM")
plt.plot(y, 'b--', label="Training Accuracy")
plt.plot(z, label="Testing Accuracy", color='orange')
plt.axis([-5,200,0,1.5])
plt.legend()
plt.grid(True)
plt.show()
```

Training and Testing Accuracies

```
C:\Users\anoos\Anaconda2\python.exe
E:/UCLA/CourseWork/Fall2017/StatisticsProgramming/Week10/Stats202A_HW9_P2.py

Training Accuracy at Iteration 0 :  0.0

Testing Accuracy at Iteration 0 :  0.934065934066

Training Accuracy at Iteration 25 :  0.992647058824

Testing Accuracy at Iteration 25 :  1.0

Training Accuracy at Iteration 50 :  0.992647058824

Testing Accuracy at Iteration 50 :  1.0

Training Accuracy at Iteration 75 :  0.992647058824

Testing Accuracy at Iteration 75 :  1.0

Training Accuracy at Iteration 100 :  0.992647058824

Testing Accuracy at Iteration 100 :  1.0

Training Accuracy at Iteration 125 :  0.992647058824

Testing Accuracy at Iteration 125 :  1.0

Training Accuracy at Iteration 150 :  0.992647058824
```

```
Testing Accuracy at Iteration 150 :  1.0
Training Accuracy at Iteration 175 :  1.0
Testing Accuracy at Iteration 175 :  1.0

Process finished with exit code 0
```



From the above graph, SVM does not overfit and provides excellent results for training and testing accuracy. Also, SVM is resilient to noise.

AdaBoost

AdaBoost is a type of "Ensemble Learning" where multiple learners are employed to build a stronger learning algorithm. AdaBoost works by choosing a base algorithm (e.g. decision trees) and iteratively improving it by accounting for the incorrectly classified examples in the training set.

The output of the other learning algorithms ('weak learners') is combined into a weighted sum that represents the final output of the boosted classifier. AdaBoost is adaptive in the sense that subsequent weak learners are tweaked in favor of those instances misclassified by previous classifiers. AdaBoost is sensitive to noisy data and outliers. In some problems it can be less susceptible to the overfitting problem than other learning algorithms. The individual learners can be weak, but as long as the performance of each one is slightly better than random guessing, the final model can be proven to converge to a strong learner.

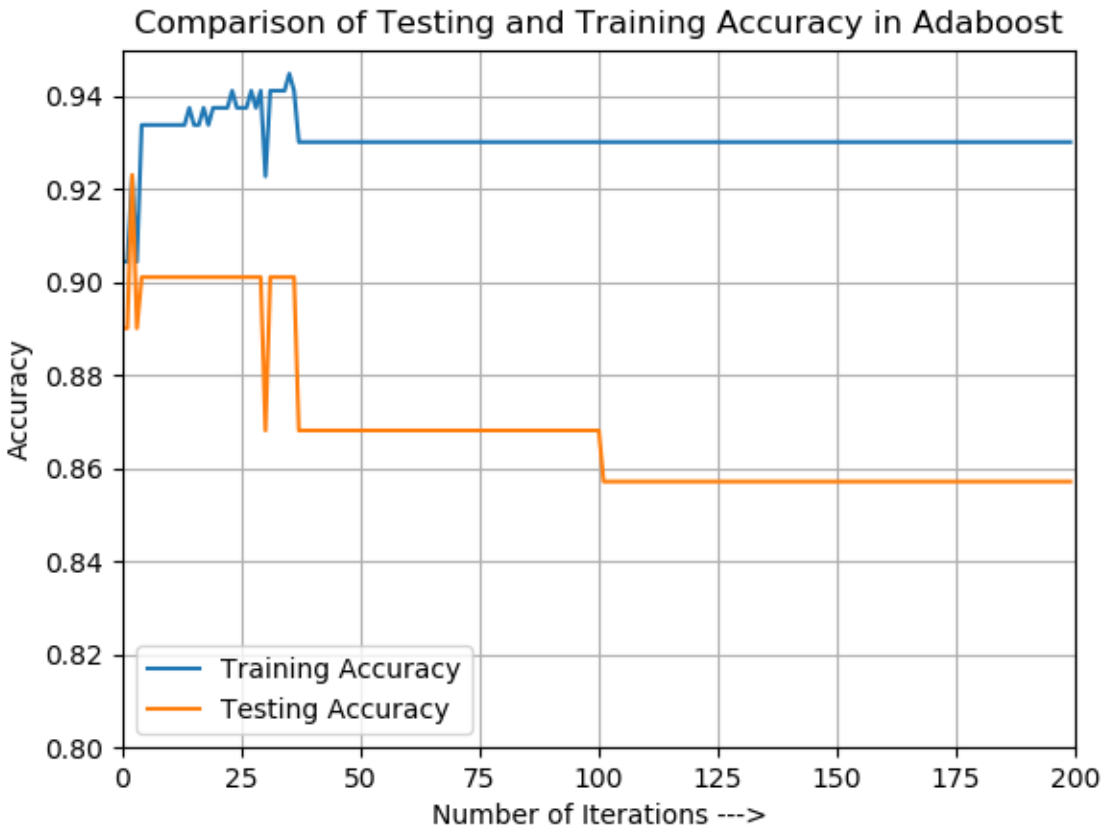
Plotting Training and Testing Accuracies

```
x, y, z = my_Adaboost(X_train, Y_train, X_test, Y_test)
plt.xlabel("Number of Iterations --->")
plt.ylabel("Accuracy")
plt.title("Comparison of Testing and Training Accuracy in Adaboost")
plt.plot(y, label="Training Accuracy")
plt.plot(z, label="Testing Accuracy")
plt.axis([0,200,0,1.5])
plt.legend()
plt.grid(True)
plt.show()
```

Training and Testing Accuracies

```
C:\Users\anoos\Anaconda2\python.exe
E:/UCLA/CourseWork/Fall2017/StatisticsProgramming/Week10/Stats202A_HW9_P2.py

Training Accuracy at Iteration  0 :  0.904411764706

Testing Accuracy at Iteration  0 :  0.89010989011

Training Accuracy at Iteration  25 :  0.9375

Testing Accuracy at Iteration  25 :  0.901098901099

Training Accuracy at Iteration  50 :  0.930147058824

Testing Accuracy at Iteration  50 :  0.868131868132

Training Accuracy at Iteration  75 :  0.930147058824

Testing Accuracy at Iteration  75 :  0.868131868132

Training Accuracy at Iteration  100 :  0.930147058824

Testing Accuracy at Iteration  100 :  0.868131868132

Training Accuracy at Iteration  125 :  0.930147058824

Testing Accuracy at Iteration  125 :  0.857142857143

Training Accuracy at Iteration  150 :  0.930147058824

Testing Accuracy at Iteration  150 :  0.857142857143
```

```
Training Accuracy at Iteration  175 :  0.930147058824
Testing Accuracy at Iteration  175 :  0.857142857143
```



AdaBoost can handle sparse datasets and therefore, can work with weak classifiers but it shows some variations after it reaches peak classification correctness. Observing the above graph, AdaBoost reaches an effective solution early but becomes sensitive to noise in further iterations.

On observing the graphs for SVM and AdaBoost, SVM provides the best results.