

Chase Headley
Neal Sakash
Dr. Anderson
CSCI 334
5/1/17

Programming Assignment Five

One thing we haven't done yet this semester was incorporating confidence intervals into our prediction models. Four different methods were used to accurately predict images from the MNIST dataset. For this assignment we built on the four models from assignment four to find the confidence interval for each. This was a fairly easy task from the sample code in the link provided. We took our accuracy scores from each method and sorted them in ascending order. For the lower bound the score at 5% of the total was determined, and for the upper bound the score at 95% was used.

```
sorted_scores = np.array(svmScores)
sorted_scores.sort()

# Computing the lower and upper bound of the 90% confidence interval
# You can change the bounds percentiles to 0.025 and 0.975 to get
# a 95% confidence interval instead.
confidence_lower = sorted_scores[int(0.05 * len(sorted_scores))]
confidence_upper = sorted_scores[int(0.95 * len(sorted_scores))]
print("Confidence interval for SVM: [{:0.3f} - {:0.3f}"].format(
    confidence_lower, confidence_upper))
```

We used this for each of the four methods:

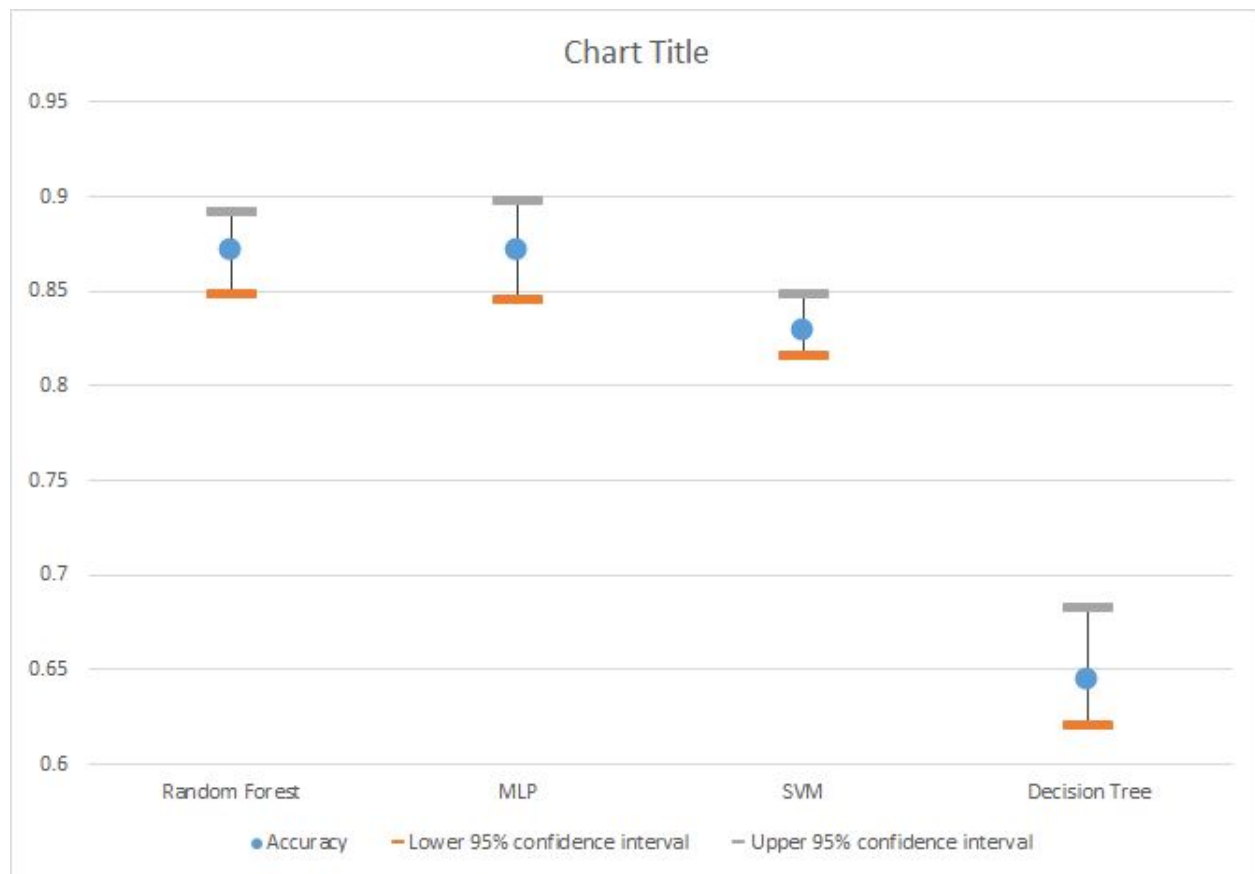
```
Random Forests: 0.871124957876
Confidence interval for Random Forests: [0.848 - 0.891]
Iteration 1: 1000 - 1.0000000000000000
```

```
MLP: 0.871155092413
Confidence interval for MLP: [0.845 - 0.897]
Iteration 1: 1000 - 1.0000000000000000
```

```
SVM: 0.829104777458
Confidence interval for SVM: [0.815 - 0.848]
Iteration 1: 1000 - 1.0000000000000000
```

```
Decision Tree: 0.644880477141
Confidence interval for Decision Tree: [0.620 - 0.682]
Iteration 1: 1000 - 1.0000000000000000
```

We can visualize the distribution of scores in the chart below:



The standard deviations within the predictions is fairly small with the largest variance occurring with the decision tree method. Including the confidence interval during analysis offers greater insight when choosing an appropriate method for modeling predictions. Not only is a high rate of accuracy sought after but also the degree of uncertainty associated with the model. In this case, even though random forest and the MLP classifier both offer a high degree of accuracy, the precision is greater and the uncertainty is smaller in the random forest model.