March 23, 2018

**Homework 3**

Sanchit Singhal

CE 395R – Data Mining with Carlos Caldas

Spring 2018

University of Texas at Austin

1. **Importance of Data Exploration in Data Mining**

Data exploration can be described as a preliminary examination of the data to better understand its characteristics. It is an important tool for data mining and there are several motivations for it. It aides in the identification of the best approach by helping to select the right tool for both preprocessing – which can sometimes be the most time-consuming task – as well, as the analysis. Apart from gaining efficiency in time, data exploration can also increase the effectiveness of the findings by helping to choose analytical methods that best suit the data and how it relates to the problem at hand. Further, it can also make use of human’s ability to inherently recognize patterns that would otherwise be very difficult to capture through data analysis tools.

1. **Neural Networks**

Advantages:

* Can provide computational efficiency through massive parallelism
* Model testing can be very fast
* Are universal approximators that can be used for a multitude of problems
* Can handle redundant attributes because weights are automatically learnt

Disadvantages:

* Model building is often very time consuming
* Can be sensitive to noise in training data
* Difficult to handle missing attributes
* Can suffer from overfitting if network is too large

1. **Accuracy as a Performance Evaluator for a classifier**

Accuracy is the most widely-used metric that tells us about the proportion of correct classifications to the total number of cases. However, a limitation of Accuracy as a metric for performance evaluation of a classifier can be that it is misleading sometimes. For example, when there are a large number of samples, n, and the model predicts everything to be a in a certain class then the accuracy score can still seem really high (99% and above) even though there are examples that it got wrong simply because the model did not detect any of the other class’s examples.