

**The University of Texas at Austin**  
**CE 395 R 5-Data Mining – Spring 2018**

**Homework 3 - Individual Assignment**

**DUE DATE: 3/23/2018**

Please submit the assignment at the beginning of the lecture.

1. Explain why data exploration is important for data mining.
2. List two advantages and two disadvantages of neural networks.
3. List one limitation of “Accuracy” as a metric used for the performance evaluation of a classifier.
4. Consider the training dataset shown in the following table. Predict the class label (stolen?) for a test example (Color=Red, Type=SUV, Origin=Domestic) using the naïve Bayes approach. Show your calculations.

Example No.	Color	Type	Origin	Stolen?
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	No
5	Yellow	Sports	Imported	Yes
6	Yellow	SUV	Imported	No
7	Yellow	SUV	Imported	Yes
8	Yellow	SUV	Domestic	No
9	Red	SUV	Imported	No
10	Red	Sports	Imported	Yes

5. Consider the training examples shown in the previous question. Which attribute would a decision tree induction algorithm choose as the first splitting attribute (root node)? Use Entropy as the measure of node impurity for selecting the best split. Show your calculations.