

CSCI 4380/6380 Data Mining

Assignment Number 3: Due 10/1/2015 (in class)

1. **[25 points] Short answers please**

- (a) Give a major difference between divide and conquer learning of decision trees and separate and conquer covering methods in data mining?
- (b) Why does a problem occur in Naive Bayes if a particular attribute value does not occur in the training set in conjunction with every class value? Briefly describe a way to fix this problem.
- (c) What is the difference between decision lists and general rule sets?

2. **[25 points][MID] Short answers please**

- (a) How does Naive Bayes handle the missing value problem in training and in testing?
- (b) How does the 1R method attempt to avoid over-fitting?
- (c) Give one advantage to using Logistic Regression over linear regression for classification.
- (d) Give one advantage for using ten-fold cross-validation over tow-fold cross-validation.

3. **[50 points]** Do exercise 17.3 on page 571 of the text book.