

CSCI 4380/6380 Data Mining

Assignment Number 3: Due 10/5/2017 (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] Short answers please

- (a) Can over-fitting happen even in the absence of noise? Briefly explain.
- (b) Why are there usually many more association rules that can be inferred from a data-set than there are classification rules?
- (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 17.6 on page 582 of the exercise handout.