

CSCI 4380/6380 Data Mining

Assignment Number 5: Due 11/17/2015 (in class)

1. [25 points] Short answers please

- (a) Give a major difference between lift charts and ROC curves.
- (b) Give one advantage to using the mean absolute error instead of the root mean squared error for performance evaluation.
- (c) Give one advantage to using the relative absolute error instead of the mean absolute error for performance evaluation.

2. [25 points] Short answers please

- (a) Why is FP-growth usually much faster than Apriori for finding large item sets?
- (b) Give one advantage to using post-pruning instead of pre-pruning for decision tree learning.
- (c) Give one advantage to using sub-tree replacement instead of sub-tree raising for decision tree pruning.