

CSCI 4380/6380 Data Mining
Second Midterm Exam - Fall 2017
Open Notes

NAME:

Problem(1):

Problem(2):

Problem(3):

Total:

1. [10 points] **Short answers please**

- (a) Give one major difference between the repeated holdout method and cross-validation for performance evaluation.
- (b) Give one major difference between cost-sensitive classification and cost-sensitive learning.
- (c) Give two major differences between the quadratic loss function and the informational loss function.
- (d) **For 6380 students only** When would the correlation coefficient be better than relative, absolute and squared errors for evaluating numeric prediction?

2. [10 points] **Short answers please**

- (a) How do support vector machines avoid substantially increasing computation time when using non linear kernels?
- (b) Give two methods for preventing over-fitting in multi-layer perceptron learning?
- (c) Give one advantage for using model trees over regression trees for continuous prediction.
- (d) Give one advantage for using regression trees over model trees for continuous prediction.

3. [10 points] Short answers please

- (a) Give two major differences between boosting and stacking.
- (b) How can boosting be applied to learners that do not use instance weights?
- (c) Give one **disadvantage** for using ensemble learning instead of single classifier learning.

