

# CSCI 4380/6380 Data Mining

## Assignment Number 4: Due 4/9/2026

1. [20 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. [20 points] Short answers please
  - (a) Give two major differences between Kernel Ridge regression and Support Vector Machines.
  - (b) Why is over-fitting less likely to happen using support vector machines than many other methods?
  - (c) Give two ways to speed up back-propagation learning of neural networks.
3. [20 points] Short answers please
  - (a) How do model trees improve upon standard regression trees, and why is smoothing used in model trees?
  - (b) Why do SVMs perform faster on sparse datasets with many zero values?
  - (c) How does the kernel perceptron differ from a standard perceptron, and what advantage does it offer?
4. [20 points] Short answers please
  - (a) Give two major difference between Bagging and Boosting.
  - (b) Give two methods for creating interpretable ensembles.
  - (c) Give one advantage of using stacking over using a single learner.
5. [20 points] Short answers please
  - a) What is the difference between scheme-independent and scheme-specific attribute selection?
  - b) Why might we discretize numeric attributes even if an algorithm can handle them directly?
  - c) What is a key difference between PCA and ICA?