STAT 5500: STATISTICAL DATA MINING

2006 Winter

Instructor: Hong Gu

Office: Chase 101

Phone: 494-7161

E-mail: hgu@mathstat.dal.ca

Lectures: 9:35-10:25am on Mon., Wed., Fri.

Place: Chase 227--seminar room

Office Hours: Mon. Fri. 11:30am-1:00pm

Course description: A variety of supervised learning and unsupervised learning methods are introduced and their statistical insights will be discussed

Prerequisite: Stat 3340.03, 4350.03, or instructor's consent

Textbook: Hastie, T., Tibshirani, R., Friedman, J. (2001),

The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer

Marking Scheme:

40 % Homework assignments

30 % Project

15 % Papers reading Report

15 % Presentation

Course Outline:

week 1-week3: Introduction, unsupervised learning methods, chapter 14.

week 4: Chapter1. Supervised learning, least square and Nearest neighbors, Sec 2.1-2.3. Statistical decision theory, Sec. 2.4.

week 5: Curse of dimensionality and Bais-variance tradeoff, 2.5-2.8. Linear methods for regression, least square and subset selection, Sec. 3.2-3.4.

week 6: Shrinkage methods, Sec. 3.4. Linear method for classification: linear discriminant analysis, Sec. 4.3

week 7: Quadratic discriminant analysis and other regularized methods, Sec. 4.3. Logistic regression, Sec. 4.4.

week 8: Generalized additive models, Sec. 9.1. CART, Sec. 9.2.

week 9: CART (continues) , Sec. 9.2. PRIM, Sec. 9.3.

week 10: MARS, Sec. 9.4. Bagging Sec. 8.7. Boosting and additive trees, Sec. 10.1-10.4.

week 11: Boosting and additive trees, Sec. 10.5-10.10.

week 12: Neural networks, Sec 11.1-11.3

week13: Neural networks.