STAT 588: DATA MINING/MACHINE LEARNING 16:960:588:01

FALL 2011, TUESDAY 6:40-9:30 PM, SEC 202 BUS

1. Course Information

- Instructor: Han Xiao
- Office: Hill Center 451
- Office Hours: Wednesday 5:00-6:00 pm or by appointment
- Email: hxiao@stat.rutgers.edu
- Text: *The Elements of Statistical Learning*, by Hastie, Tibshirani and Friedman. Springer, 2009, 2ed. Full text available from Springer `http://dx.doi.org/10.1007/978-0-387-84858-7` Access from campus or login via Rutgers account. You may also visit the website of the book: `http://www-stat.stanford.edu/ tibs/ElemStatLearn/`.
- Software: R. Free software available at `http://www.r-project.org/`. If you go to `Manuals` on the left panel of the website, you will find a good introduction *An Introduction to R*. A more advanced reference is *Modern Applied Statistics with S*, by Venables and Ripley. Springer, 2002, 4ed.
- Course website: `http://stat.rutgers.edu/home/hxiao/`
- Course work: four homework assignments plus one take home final exam.
- Grades: homework (70%), final exam (30%).

2. Syllabus (tentative)

Week #	Date	Topic	Reading	HW[1]
2	Sep 06	Introduction	Ch. 1+2	
3	Sep 13	Training versus testing and linear regression	Ch. 3	HW1 A
4	Sep 20	Regularization and variable selection	Ch. 3	
5	Sep 27	Unsupervised learning (I) - PCA	Ch. 4+14	
6	Oct 04	LDA and linear classification	Ch. 4	HW1 D, HW2 A
7	Oct 11	Binary classification	Ch. 4	
8	Oct 18	Basis expansion	Ch. 5	
9	Oct 25	Kernel methods	Ch. 6	HW2 D, HW3 A
10	Nov 01	Model assessment and selection	Ch. 7	
11	Nov 08	Unsupervised learning (II) - clustering	Ch. 13+14	
12	Nov 15	Unsupervised learning (III) - other methods	Ch. 14	HW3 D, HW4 A
13	Nov 22	Model averaging	Ch. 8	
14	Nov 29	Decision tree and boosting	Ch. 9+10	
15	Dec 06	Neural networks	Ch. 11	HW4 D, Final A
16	Dec 13	SVM and high dimensional problems	Ch. 12+18	
17	Dec 20			Final D

[1] A=Assigned, D=Due.