EXST7152: Advanced Topics in Statistical Modeling

Instructors: Dr. Bin Li
Offices: Office: Rm 173 Woodin Hall, LSU
Contacts: 578-1343 (bli@lsu.edu)

Reference: *The Elements of Statistical Learning - Data Mining, Inference and Prediction*, Hastie, T., Tibshirani, R., and Friedman, J. (2nd Ed., 2009), New York: Springer.
An Introduction to Statistical Learning with Applications in R, James, G., Witten, D., Hastie, T. and Tibshirani, R. (1st Ed. 2013), Springer.

Handouts: The handouts can be purchased in LSU Barnes & Noble Bookstore. The address is 2 Union Square, Baton Rouge, LA 70803. The telephone number is (225) 578-51370.

Prerequisites: EXST 7013/7014/7015 and EXST 7034 or equivalent or consent of department head. Familiarity with matrix algebra and linear regression analysis, and working knowledge of one or more standard software packages such as `R`, `S-plus`, `SAS`, `C/C++` or `Matlab`.

Course description:
Contemporary statistical methodology explores various ways of estimating functional dependencies between a response variable and possibly a large set of features, when one is trying to find and understand the unknown underlying mechanism within the realm of noisy and/or complex data. Modern regression and pattern recognition analyses fall in this framework. This course will provide an overview of modern statistical methodology in regression and pattern recognition problems such as regularized linear regression and classification methods, penalized splines and signal regression analysis, tree-based methods, support vector machines and kernel-related methods, and boosting and illustrate the practical uses of the algorithms.

Course web: `http://statweb.lsu.edu/faculty/li/teach/exst7152/`

Computing: The lecture examples will be illustrated using `R`, which is a free software environment for statistical computing and graphics. No restrictions on softwares used in project.

Final Score: Homework (35%), team project and presentation (20%)
(45%) Final project and final exam.

List of topics:

Introduction to statistical modeling & `R`
Introduction to resampling methods
      Introduction to Bootstrapping
      Resampling methods for spatial problems
Model evaluation
Regularized linear regression methods
      Shrinkage methods: ridge and lasso
      Lasso-related methods
Regularized linear classification
      Flexible and penalized discriminant analysis
      L_1 regularized logistic regression
Tree and related methods
      Classification and regression tree (CART)
      Multivariate additive regression splines (MARS)
      Patient rule induction method (PRIM)
Ensemble methods and model averaging
      Bagging
      Random forest
      Boosting
Support vector machines and kernel-related methods
      Support vector machines (SVM) and kernels
      Kernel-related methods

Readings

1. "Statistical Modeling: The Two Cultures", Leo Breiman. (2001) *Statistical Science.* **16**, 199-231.

2. "Classifier Technology and the Illusion of Progress", David J. Hand, *Statistical Science.* **21**, (2006), 1-14.

3. "Regularization and variable selection via the elastic net", Zou and Hastie, (2005) *Journal of the Royal Statistical Society. Series B.* **67**, 301-320.

4. "Least Angle Regression", B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, *Annals of Statistics* (2004), **32** 407-451.

5. "Sparsity and smoothness via the fused lasso", Tibshirani and et al. (2005) *Journal of the Royal Statistical Society. Series B.* **67**, 91-108.

6. "Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy", Efron, B. and Tibshirani, R. (1986), *Statistical Science*, **1**, 54-75.

7. Wolpert, David H., 2001. "The supervised learning no-free-lunch theorems". *Proceedings of the 6th Online World Conference on Soft Computing in Industrial Applications.*

8. Friedman, J. H. and Popescu, B. E. "Importance Sampled Learning Ensembles." (2003)

9. Friedman, J. H. and Popescu, B. E. "Gradient directed regularization." (2004).

10. Friedman, J. H. and Popescu, B. E. "Predictive Learning via Rule Ensembles." (2005).

11. "Kernels and Ensembles: Perspectives on Statistical Learning", Mu Zhu. (2008) *The American Statistician*, **62**, 97-109.