

**Homework 3**  
**Due Wednesday [05/27/13] in APM 5151 3pm**  
**MATH 287D – Statistical Learning**

*Be concise. Always comment on your results/findings. Send your R code, cleaned, polished and commented. Make it so that running the R code is straightforward. Send it to [jbradic@ucsd.edu](mailto:jbradic@ucsd.edu) with the exact subject line 287D Homework (number).*

**Problem 1.** The following is called the probit regression model. Suppose  $Y \in \{0, 1\}$  is a random variable given by

$$Y = \begin{cases} 1, & a^T X + b + V \leq 0 \\ 0, & a^T X + b + V > 0 \end{cases}$$

where  $X \in R^p$  is a vector of explanatory variables and  $V \sim \mathcal{N}(0, 1)$  is a latent (unobserved) random variable. Formulate the maximum likelihood estimation problem of estimating  $a$  and  $b$ , given data consisting of pairs  $(X_i, Y_i)$  as a convex optimization problem.

**Problem 2.** In this problem you will fit a logistic regression model to the UCI Pima Indians diabetes database. The data, and a description of the data, can be downloaded from <http://archive.ics.uci.edu/ml/datasets/Pima+Indians+Diabetes>. It is a binary classification problem, with 768 instances having eight features each.

- (a) Fit a maximum likelihood logistic regression model using Newton's method (iteratively reweighted least squares). Summarize your findings.
- (b) Now use sparse logistic regression: add Lasso penalty on logistic loss function. Plot the estimates as a function of  $\lambda$ . Also, plot the AIC score as a function of  $\lambda$ .
- (c) Apply SVM classifier to the same dataset. Explain your method and implementation.
- (d) Apply sparse SVM classifier to the same dataset.
- (e) Use theory done in class to show risk bounds for logistic and for SVM regression method. Identify what is  $\phi$  and  $\Psi$  for each classification method.

**Problem 3.**

1. Read the paper *Convexity, Classification, and Risk Bounds* by Peter L. Bartlett, Michael I. Jordan and Jon D. McAuliffe  
<http://statistics.berkeley.edu/sites/default/files/tech-reports/638.pdf>.