Homework 2
Due Friday [05/01/13] 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 with the exact subject line 287D Homework (number).

Problem 1.
Extend simple lemma for Lasso model selection property (from class) to the case of more dimensions.

1. I.e. prove that Lasso "almost never" selects the correct model $\beta^* = (\beta_1^*, \beta_2^*, \cdots, \beta_s^*, 0, 0, \cdots, p)$, even in case of orthonormal designs, when the tuning parameter is selected to minimize the prediction error.

2. In R, simulate each model in Problem 1. Varying the sample size (e.g., $n \in \{50, 200, 1,000\}$), compute the Lasso coefficients (see the function `lars()`) and show that they do not converge to the true values.

Problem 2.

1. Read the paper *Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties* } *available from* `http://orfe.princeton.edu/~jqfan/papers/01/penlike.pdf`

2. *Compute coefficients of SCAD penalty for the case of orthonormal designs i.e. when* $\mathbf{X}^T\mathbf{X} = \mathbf{I}$.

3. *Download* `Lars` *and* `ncvreg` *packages in R and familiarize yourself with them (focus on understanding the outputs). Propose how to compare SCAD and Lasso using simulations. Perform the analysis. Use graphical outputs as well.*

Problem 3.

1. Read the paper *Simultaneous analysis of Lasso and Dantzig Selector* The Annals of Statistics 2009, Vol. 37, No. 4, 1705–1732

2. Show using simulations how correlation might effect Lasso performance. Will the same correlation ruin Scad estimator?

Problem 4. (Bonus) Derive KKT conditions for the following problem

$$\hat{\beta} = \arg\min \{\|\Sigma\beta - \delta\|_\infty \leq \lambda_n + \gamma_n \|\beta\|_1\}$$

where $\Sigma \in R^{p\times p}$ and $\delta \in R^p$ are known matrix and vector.