ST790-003, Homework 5 (Updated Mar 1)

Due Friday, Mar 20, 2015 @ 11:59PM

QP and SOCP

This homework explores applications of quadratic programming (QP) and second order cone programming (SOCP) in statistics. This is a *solo* homework. Discussion with fellow students is allowed but you have to write your code and report independently.

1. (Lasso regression) We again work on the prostate cancer data in HW4.

   (a) Repeat HW4 Q1 parts (a) and (b).

   (b) Fit lasso regression

   $$\text{minimize} \quad \frac{1}{2}\|\boldsymbol{y} - \beta_0\mathbf{1} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1$$

   at $\lambda = 0, 2, 4, \ldots, 64$ on the training data and plot solution path. Also plot the prediction errors $\sum_{i \in I_{\text{test}}}(y_i - \hat{y}_i)^2/n_{\text{test}}$ on the test set over λ.

2. (Group lasso) We again work on the prostate cancer data in HW4.

   (a) Repeat HW4 Q1 parts (a) and (b).

   (b) We suspect some nonlinear effects of predictors on the response variable and thus add quadratic and cubic terms of all (centered and scaled) continuous predictors (`lcavol`, `weight`, `age`, `lbph`, `lcl`, `pgg45`) into the model.

   (c) Fit group lasso regression

   $$\text{minimize} \quad \frac{1}{2}\|\boldsymbol{y} - \beta_0\mathbf{1} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2 + \lambda\sum_{g=1}^{G}\sqrt{p_g}\|\boldsymbol{\beta}_g\|_2$$

   at $\lambda = 2^{-3}, 2^{-2.8}, 2^{-2.6}, \ldots, 2^{4.8}, 2^5$ on training data and plot solution path. Here each group g corresponds to all polynomial terms of one predictor and p_g is group size. That is $p_g = 1$ for categorial variables `svi` and `gleason` and $p_g = 3$ for continuous variables. Also plot the prediction errors on the test set over λ. In both plots, use log scale for the x-axis.

3. (SVM) We again work on the South African heart disease data in HW4.

   (a) Repeat HW4 Q6 part (a).

   (b) Fit the support vector machine

   $$\text{minimize} \quad \sum_{i=1}^{n}\left[1 - y_i\left(\beta_0 + \sum_{j=1}^{p}x_{ij}\beta_j\right)\right]_+ + \lambda\|\boldsymbol{\beta}\|_2^2$$

at $\lambda = 0, 2, 4, \ldots, 100$ on the training data and plot solution path. Also plot the misclassification rates on the test data over λ. Compare the results to those you obtained in HW4 Q6.

4. (Denoise Lena) A picture of Lena (`lena256noisy.png`) is corrupted by noise. Your goal is to recover a cleaner picture by denoising.

   (a) Let $\boldsymbol{y} = (y_{ij}) \in \mathbf{R}^{m \times n}$ represents the gray levels across a 2D array of pixels from a noisy image with true gray levels $\boldsymbol{x} = (x_{ij})$. The ROF model (Rudin et al., 1992) minimizes the total variation regularized least squares criterion

$$\frac{1}{2}\|\boldsymbol{y} - \boldsymbol{x}\|_{\mathrm{F}}^2 + \lambda \sum_{i,j} \sqrt{(x_{i+1,j} - x_{ij})^2 + (x_{i,j+1} - x_{ij})^2}. \tag{1}$$

   The total variation penalty serves to smooth the reconstructed image and preserve its edges. Discuss how you can optimize the objective (1) as an SOCP. You do *not* have to implement it.

   (b) A similar effect can be achieved by the anisotropic penalty

$$\frac{1}{2}\|\boldsymbol{y} - \boldsymbol{x}\|_{\mathrm{F}}^2 + \lambda \sum_{i,j} \left(|x_{i+1,j} - x_{ij}| + |x_{i,j+1} - x_{ij}|\right). \tag{2}$$

   Minimize the objective function (2) at $\lambda = 5, 15, 50$ and display the recovered images.

References

Rudin, L. I., Osher, S., and Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1-4):259 – 268.