Homework 3

March 21, 2018

Data Analysis

This exercise is to learn to use Lasso, Ridge regression, PCR, PLS, and as a byproduct cross-validation. Data should be used to examine the relation between the level of prostate-specific antigen lpsa and a number of clinical measures in men who were about to receive a radical prostatectomy: lcavol (log cancer volume), lweight (log prostate weight), age (in years), lbph (log of the amount of benign prostatic hyperplasia), svi (seminal vesicle invasion), lcp (log of capsular penetration), gleason (a numeric vector), pgg45 (percent of Gleason score 4 or 5).

Finally the data set cointains a column called train a logical vector. You should consider the rows for which train=T and use them as your training set and run on it Lasso, RR, PCR, OLS. Use the rest of the data as a validation set, to estimate prediction error.

The exercise consists in reproducing Figure 3.7, and Table 3.3 of ELS, and Figure 3.8, 3.10. Please submit only the tables and plots you obtain, and an ASCII file with the code you use (and please since there is randomness in this exercise, you should fix the seed set.seed(a number) so that your exercise can be reproducible).

Theoretical Questions

The first two questions consist basically in filling some (temporary) omissions in the third set of notes. See the notes3.pdf file, where I highlighted where the missing parts should be inserted. To answer the third question, you may want to look at page 4 on the notes2.pdf file, the **Another Example** section.

1) Compute the mean sqaured error for ridge regression in the coordinate system of the eigen directions of the predictor space (for simplicity just assume the case of maximum rank), that is, where the matrix $X^T X$ is diagonal $\mathrm{diag}(d_1^2, \ldots, d_p^2)$. Assume that the training predictors are not random (that is, only average over the error distribution). In particular show that the variance of the estimator $\hat{F}(\mathbf{x}) = \mathbf{x}^{\mathbf{T}}\hat{\beta}$ at the point $\mathbf{x}$ is

given by

$$Var(\mathbf{x}) = \sigma^2 \sum_{i=1}^{p} x_i^2 \frac{1}{d_i^2} \frac{d_i^4}{(d_i^2+\lambda)^2}$$

and the bias is given by

$$B(\mathbf{x}) = \sum_{i=1}^{p} x_i \beta_i \left(\frac{d_i^2}{d_i^2+\lambda} - 1 \right)$$

(You can ignore $\hat{\beta}_0$ by assuming Y was been centered).

2) Show that in the case of orthogonal regressors $X^T X = I_p$ the Lasso estimates are given by

$$\hat{\beta}_i = \text{sign}(\hat{\beta}_i^{OLS}) \cdot \max\{|\hat{\beta}_i^{OLS}| - \lambda, 0\} = \text{sign}(\hat{\beta}_i^{OLS}) \left(|\hat{\beta}_i^{OLS}| - \lambda \right)_+$$

where $(x)_+$ is the positive part of x.

3) This exercise is a similar case to the worked-out example in the second set of lectures (page 4). The only difference is that now you have to consider other losses. Just choose one of the two losses only.

Consider the training set $T = \{(y_i, x_i)\}_{i=1,\ldots,n}$, where for simplicity we assume x_i to be fixed and $y_i \in \{-1, 1\}$ to be a binary random variable with $p_i = P(y_i = 1)$. Given a training set, we assume our estimate for each p_i will be $\hat{p}_i = (1 + a y_i)/2$, where $0 \leq a \leq 1$ is a parameter that controls the degree of fit to the training data. Larger values provide a closer fit. We want to compute training and test error using one of the following two losses: exponential

$$L_1(y, F) = \exp(-yF)$$

and the squared error

$$L_2(y, F) = (y - F)^2$$

a) For the loss of your choice, F is defined as the population minimizer of the coresponding population risk:

$$F = \arg\min_F E[L(y, F)]$$

Show that

$$F_1 = \frac{1}{2} \log\left(p/(1-p)\right)$$

using L_1 and

$$F_2 = 2p - 1$$

using L_2.

b) Show now that the training error $\hat{R}$ (the average loss on the training data) and the test error R (average population risk) are

$$\hat{R}_1 = \left(\frac{1-a}{1+a}\right)^{1/2} \qquad R_1 = (1-\bar{e})\left(\frac{1-a}{1+a}\right)^{1/2} + \bar{e}\left(\frac{1+a}{1-a}\right)^{1/2}$$

and

$$\hat{R}_2 = (1-a)^2 \qquad R_2 = (1-a)^2 + 4\bar{e}a$$

where $\bar{e} = \frac{2}{N}\sum_{i=1}^{N} p_i(1-p_i)$