BIOS:7600 Homework 6

Chuan Lu

April 6, 2019

1. Problem 4.2

   (a) *Proof.* By (4.3) in the book, since both $\beta_j, \beta_k > 0$,

   $$x_j^\top (y - X\hat{\beta})/n - \lambda_2 \hat{\beta}_j = \lambda_1,$$
   $$x_k^\top (y - X\hat{\beta})/n - \lambda_2 \hat{\beta}_k = \lambda_1.$$

   Then

   $$(x_j - x_k)^\top (y - X\hat{\beta})/n - \lambda_2(\hat{\beta}_j - \hat{\beta}_k) = 0,$$

   thus

   $$\beta_j - \beta_k = \frac{1}{n\lambda_2}(x_j - x_k)^\top r.$$

   □

   (b) *Proof.*

   $$\begin{aligned} |\beta_j - \beta_k| &= \frac{1}{n\lambda_2}|(x_j - x_k)^\top r| \\ &\leq \frac{1}{n\lambda_2}\|(x_j - x_k)\| \cdot \|r\| \\ &= \frac{1}{n\lambda_2}\sqrt{2n(1-\rho)}\|r\|. \end{aligned}$$

   Since $\hat{\beta}$ is the unique minimizer of the loss, so by substituting β to 0,

   $$\|r\| \leq \|y\|.$$

   Then we have shown the result. □

2. Problem 4.5

   The code for this problem is `problem4.5.R`.

   (a) Use elastic net with cross validation to fit the unfiltered data. The result of R^2 and degree of freedom is shown in Table 2a.

   (b) The number of variables after filtering is 857. The result of using again elastic net with cross validation is shown in Table 2b.

α	$\hat{R^2}$	df
1	0.9726997	82
0.75	0.966037	80
0.5	0.9757089	108
0.25	0.9468617	116

Table 1: R^2 and *df* for different α for unfiltered data.

α	$\hat{R^2}$	df
1	0.907334	46
0.75	0.9795399	94
0.5	0.97745	109
0.25	0.9684986	126

Table 2: R^2 and *df* for different α for filtered data.

(c) i. On *df*. By filtering out certain features, one can expect a decrease in the number of variable selected, especially when α is large, i.e., when including more Lasso penalty. However, this may exclude some important features from the regression model.

ii. On accuracy. If we filter by biological annotation, the predictive accuracy is almost the same with the unfiltered case. However, if we filter by variance, the accuracy is greatly less than the unfiltered version. Hence, when filtering before analysis, the accuracy may decrease, mainly because of the previous argument, i.e., some important features are excluded.