ST790 Homework 3 Solution

1.

The objective function of the elastic net problem

$$\begin{aligned} L_{enet}(\lambda_1, \lambda_2, \beta) &= \|y - X\beta\|_2^2 + \lambda_2\|\beta\|_2^2 + \lambda_1\|\beta\|_1 \\ &= \left\| \begin{pmatrix} y \\ 0 \end{pmatrix} - \begin{pmatrix} X \\ \sqrt{\lambda_2} I_p \end{pmatrix} \beta \right\|_2^2 + \lambda_1\|\beta\|_1 \end{aligned}$$

is equivalent to the objective function of the lasso problem with augmented data $X^* = (1 + \lambda_2)^{-1/2}(X^T, \sqrt{\lambda_2} I_p)^T$ and $y^* = (y^T, 0^T)^T$,

$$L_{lasso}(\gamma, \beta^*) = \|y^* - X^*\beta^*\|_2^2 + \gamma\|\beta^*\|_1,$$

where $\gamma = \lambda_1/\sqrt{1 + \lambda_2}$ and $\beta^* = \sqrt{1 + \lambda_2}\beta$. The minimizer of L_{lasso}, $\hat{\beta}^*$, can be transformed back to the minimizer of L_{enet} using $\hat{\beta} = \hat{\beta}^*/\sqrt{1 + \lambda_2}$. Here we devide the design matrix by $\sqrt{1 + \lambda_2}$ to make sure that each column of X^* has sum of squares $\sum_{i=1}^{n+p} x_{ij}^{*2} = 1$.

2.

(a) According to the `summary()` output, the R^2 is 0.6944, and the set of significant predictors (at the level $\alpha = 0.05$) is {lcavol, lweight, lbph, svi}. The $TrainErr$ is around 0.4392, and the $TestErr$ is around 0.5213.

(b) Using the BIC, model $\widehat{M_2}$ is selected, its set of significant predictors is {lcavol, lweight}. If we refit the OLS using the selected predictors, the $TestErr$ is around 0.4925.

(c) Using the AIC, model $\widehat{M_7}$ is selected, its set of significant predictors is {lcavol, lweight, age, lbph, svi, lcp, pgg45}. If we refit the OLS using the selected predictors, the $TestErr$ is around 0.5165.

3.

(a) Since we select the parameter with `cv.lars(x, y, K = 5)`, the results for this question depend on how the training set are shuffled into the five folds. Using the minimum CV rule, the best fraction could be around 0.5 or 0.9. Using the one-standard deviation rule, the best fraction could be around 0.35.

(b) The value of the best λ according to the BIC may vary as the training data are preprocessed differently (whether `x` and `y` are standardized, whether the intercept is included). But the selected model should have the significant predictors {lcavol, lweight, lbph, svi, pgg45}. The $TestErr$ is around 0.46.

(c) The best fraction also varies but should be around 0.3 according to the BIC. The selected model should have the significant predictors {lcavol, lweight, svi}. The $TestErr$ is around 0.44.