BIOS:7600 Homework 4

Chuan Lu

March 4, 2019

1. Problem 2.3

   Proof. By 2.2(b), we know $L(\eta|y)$ is a strictly convex function of η. Notice $\|\beta\|_1$ is a convex function of β, and since there is an affine mapping between β and η, we know $Q(\eta) = Q(\beta|X, y)$ is strictly convex w.r.t. η.

   By the property of strictly convex functions, Q only has one minimizer, which means $X\beta_1 = X\beta_2$. □

2. Problem 2.6

   The code is submitted by Dropbox (`lasso.R` and `test.R`).

3. Problem 2.7

   (a) $\lambda = 0.2335721$ is chosen by 10-fold cross validation. The code for cross validation is submitted by Dropbox (`cross_validation.R`).

   The number of nonzero coefficients is 9. We check that lasso is better by a seperate test (since for this model, the whole dataset is used for fitting and thus we do not know out-of-sample prediction errors). 500 points are used for fitting and the rest for test. The MSPE for lasso is 17.14, for OLS is 23.64, for null model is 17.10. So there is evidence that this model is better than OLS, but no such evidence that it is better than the null model.

   (b) The most important variables by lasso are (not ranked by importance):
      - age
      - rr
      - hfa
      - inc
      - lcw
      - aro
      - qcr
      - afe

- absu

(c) The number of nonzero coefficients is significantly smaller than the total number of variables. The result of p-values of the OLS model shows there are 14 siginficant coefficients, hence OLS is more liberal than lasso.

(d) The result is as follows, where Lasso and ridge estimates are similar while the OLS estimate is different. By the same reason that these two abilities are similar, I think both lasso and ridge estimates are more reasonable than OLS.

	absu	afe
lasso	0.0276	0.0662
ridge	0.0358	0.0896
OLS	-0.0270	0.1531

4. Problem 2.8 (code: `problem2.8.R`)

(a) The results are as follows. Since there are only two significant variables, Lasso is expected to have the most accurate estimate.

forward	ridge	lasso
0.742	0.625	0.203

(b) There's an error "AIC is -infinity for this model, so step cannot proceed". However, it is more extreme than the previous simulation, so lasso would also have the most accurate estimate, and the difference of lasso with other two methods would be larger than the previous one.

(c) The result is as follows. Lasso still does the best job, but the difference is smaller; It's because the number of significant variables increases while the absolute value of coefficient of each variable decreases.

forward	ridge	lasso
0.920	0.637	0.585

(d) The same error as in (b). However, I expect ridge regression and forward selection might have the same error with lasso. There's no strong sparsity in this model, so lasso could not do a better job.