Stat 432 Homework 6

Assigned: Mar 3, 2019; Due: Mar 8, 2019

Question 1 (ridge regression) [5 points]

On page 27 of the lecture note `Penalized`, we introduced a formula for calculating the degrees of freedom of a ridge regression. From the previous kNN model, we also know that the degree of freedom is N/k. Use the Boston housing data again for this analysis. Let's use a cross-validation approach to evaluate the performance of the ridge regression and kNN for a grid of degrees of freedoms. To do this, first, choose a sequence k values such that the degrees of freedom is within the range of 1 to 14. The choose the corresponding λ values in the ridge regression such that the degrees of freedom of these two models are matched. Then use a 10-fold cross-validation to evaluate and compare their performance. You can use `caret` package, as well as functions such as `lm.ridge`, `knn` or any build in functions to perform the model fitting. You should consider using plots to demonstrate the results.

Note: You should be careful about three things: 1) how scaling in ridge affect the degrees of freedom; 2) how does kNN take care of categorical variables. 3) Intercept in the ridge regression is not penalized, so there is 1 df for that. Explain or justify your approach.

```r
data(Boston, package="MASS")
# head(Boston)

useLog = c(1,3,5,6,8,9,10,14)
Boston[,useLog] = log(Boston[,useLog])
Boston[,2] = Boston[,2] / 10
Boston[,7] = Boston[,7]^2.5 / 10^4
Boston[,11] = exp(0.4 * Boston[,11])/1000
Boston[,12] = Boston[,12] / 100
Boston[,13] = sqrt(Boston[,13])
```

Question 2 (Lasso regression) [5 points]

Use the Boston housing data again to perform the Lasso regression. For this question, you should consider using the `glmnet` and the corresponding cross-validation version `cv.glment` to tune the parameters. Perform a complete Lasso regression analysis of this data, such as plotting the cross-validation errors, and how the estimated parameters change as a function of λ. Select the best tuning that minimizes the cross-validation error and report the selected variables. Compare this result to the best subset selection with AIC penalty. Based on what we have learned, comment on how these two methods trade the bias-variance differently.

Extra-Credit Question [4 points]

On pages 8-9 of lecture 5.1, we discussed the coin tossing example that violates the strong likelihood principle. The scientist tosses the coin 12 times, of which only 3 are heads. The statistician says: "you tossed the coin 12 times and you got 3 heads. The one-sided p-value is 0.07". Then the scientist says: "Well, it wasn't exactly like that… I actually repeated the coin tossing experiment until I got 3 heads and then I stopped". The statistician say: "In that case, your p-value is 0.03". Given explanations for the two different conclusions.