

Homework 10

Due Tuesday, December 10

Show all your work. Data files available from `alr4` or `MASS` packages, or on class website. (NOTE: In order to use the `glmnet` package, you will need to convert the data to numeric, just as it was done in `longley` dataset).

1. Using the `longley` dataset in R, regress `Employed` on all other variables by LASSO regression. For the LASSO regression, use `glmnet()` and show the plot, and perform `cv.glmnet()` and show the plot, and show the coefficients using `lambda.min` and `lambda.1se`. Compare the selected coefficients with those of linear regression variable selection (`step`). Comment.
2. Refer to Problem 1. Perform the ridge regression. Again, use `glmnet()` and `cv.glmnet()` as in Problem 1, and compare the coefficients with those of linear regression. Comment.
3. Recall HW 8, Problem 3. Perform the LASSO regression (with the same steps as outlined in Problem 1) and compare the result from the previous homework. Interpret and comment.
4. Recall HW 8, Problem 4. Perform the LASSO regression (with the same steps as outlined in Problem 1) and compare the result from the previous homework. Interpret and comment.
5. Please find the dataset `violshelt.txt`.

Variable Description

<code>place</code>	Placement (0: Outpatient, 1: Residential)
<code>age</code>	Age at Admission (Years)
<code>race</code>	Race (1: White, 2: Non-white)
<code>gender</code>	Gender (1: Female, 2: Male)
<code>neuro</code>	Neuropsychiatric Disturbance (1: None, 2: Mild, 3: Moderate, 4: Severe)
<code>emot</code>	Emotional Disturbance (1: Not Severe, 2: Severe)
<code>danger</code>	Danger to Others (1: Unlikely, 2: Possible, 3: Probable, 4: Likely)
<code>elope</code>	Elopement Risk (1: No Risk, 2: At Risk)
<code>los</code>	Length of Hospitalization (Days)
<code>behav</code>	Behavioral Symptoms Score (0 - 9)
<code>custd</code>	State Custody (1: No, 2: Yes)
<code>viol</code>	History of Violence (1: No, 2: Yes)

Using `place` as a response, fit a logistic regression and perform a stepwise variable selection. Then fit a LASSO logistic regression model (with the same steps as outlined in Problem 1). Compare the results and comment.

6. Please find the dataset `gxp.txt`, a gene expression data. Using `y` (labeled 0 or 1) as the response and `x1` to `x100` (genes) as predictors, fit an elastic net (with `alpha=0.5`) and a LASSO logistic regressions. In both cases, follow the same steps as outlined in Problem 1. Compare the results and comment.