

Machine Learning 201

Homework 2

Mike Bowles, PhD & Patricia Hoffman, PhD

Look at the [elasticnet.pdf](#) (which has been uploaded to the Machine Learning 201 web site:

<http://machinelearning201.pbworks.com/f/elasticnet.pdf>) The elasticnet r package describes the diabetes data. This homework uses x_2 , the matrix with 64 columns and the numeric target vector y presented on page 4 of that pdf. At the bottom of this page is some r code that will help you obtain this data.

- 1) Use LAR on the diabetes data set with 10 fold cross validation. What is the least squared error? Which variables are included in the optimum solution?
- 2) Next use lasso on the diabetes data set 10 fold cross validation. What is the least squared error? Which variables are included in the optimum solution?
- 3) Now use ridge regression on the diabetes data set 10 fold cross validation. What is the least squared error? Which variables are included in the optimum solution? Was there an improvement?
- 4) Finally use elastic net on the data set with 10 fold cross validation.
- 5) Plot the coefficients obtained by using these techniques on the same graph. The horizontal axis should be the 64 types of observations and the vertical axis should be the coefficients obtained for these observations for each of the 4 techniques that were used on this data. Compare and contrast the solutions from the first four problems. Are the coefficients of the factors, which were eliminated with lasso, small after applying ridge regression? Are there other interesting observations you can make?

```
install.packages("elasticnet")
help(package="elasticnet")
require("elasticnet")
data(diabetes)
```

attach(diabetes)