

# Homework 2, STATS 315A

Stanford University, Winter 2019

*Joe Higgins*

## Question 2

This question will have you write an R function to efficiently perform forward stepwise linear regression. It requires that you understand the material on pages 13–15 of the chapter 3 lecture notes. The setup is linear regression with an  $N \times p$  matrix  $X$  and a response vector  $y$ . You always include an intercept in your models, which is not included in  $X$ .

- (a) Let  $\tilde{X}$  be the matrix  $X$  with each of the columns mean centered. What is the fitted intercept in the regression of  $y$  on  $\tilde{X}$ ? How do the coefficients of  $X$  in the regression of  $y$  on  $X$  compare with the coefficients of  $\tilde{X}$ ?

```
#stuff goes here
```

## Question 7

7. Obtain the zipcode train and test data from the ESL website.
- Compare the test performance of a) linear regression b) linear discriminant analysis and c) multiclass linear logistic regression.
  - For a) and c), use the package glmnet (available in R, matlab and python) to run elastic-net regularized versions of each (use  $\alpha = 0.3$ ). For these two, plot the test error as a functions of the training  $R^2$  for a) and  $D^2$  for c) (% training deviance explained).
  - In ii., what is the optimization problem being solved?

```
train <- read.csv("zip.train")
test  <- read.csv("zip.test")
```