

Homework_2

Lindley Slipetz

7/11/2021

For this homework, I will be using the Childhood adversity and traumatic stress among inpatients at a psychiatric hospital in the Baltimore area from 1993-1995. The data include diagnoses, psychological symptoms, physical and sexual abuse, post-traumatic stress disorder, self-destructive behavior, and demographic data. I will be predicting psychoticism from gender, SES, age, occurrence of mood disorder, paranoid ideation, and level of substance abuse.

Let's load the data and packages!

```
library(caret)
#install.packages("glmnet")
library(glmnet)
library(pROC)
library(tidyverse)
#install.packages("e1071")
library(e1071)
full_data <- read.table(file = 'G:\\My Drive\\ICPSR\\ML\\HW_2\\36168-0001-Data.tsv', sep = '\\t', header
```

Okay, let's subset our data to just the variables we're interested in.

```
subset_data <- full_data %>%
  select(PSYCDX, SEX, SES, MOODDX, SCL_PAR, SISDB_SUB, AGE)
```

Now we're going to check if there's any missing data.

```
df <- as.data.frame(
  cbind(
    lapply(
      lapply(subset_data, is.na), sum)
    )
  )

rownames(subset(df, df$V1 != 0))
```

```
## [1] "SCL_PAR" "SISDB_SUB"
```

Okay, SCL_PAR, and SISDB_SUB have missing values. Let's see how much of a problem this is.

```
sum(is.na(subset_data$SCL_PAR))
```

```
## [1] 1
```

```
sum(is.na(subset_data$SISDB_SUB))
```

```
## [1] 10
```

That's not that much missing data (at least to me). I think we'd be safe to just omit the data with NA.

```
complete_data <- na.omit(subset_data)
```

Dummy encoding.

```
complete_data$PSYCDX <- factor(complete_data$PSYCDX, labels=c("non_psy", "psy"))
createDummies <- dummyVars(~., complete_data[, -1], fullRank = TRUE)
new.predictors <- predict(createDummies, complete_data[, -1])
complete_data <- data.frame(PSYCDX = complete_data$PSYCDX, new.predictors)
```

Time to standardize the data.

```
preProcValues <- preProcess(complete_data, method=c("center", "scale"))
complete_data <- predict(preProcValues, complete_data)
```

Splitting the data.

```
set.seed(1985)
trainIndex <- createDataPartition(complete_data$PSYCDX, p=0.7, list=FALSE)
train <- complete_data[trainIndex,]
test <- complete_data[-trainIndex,]
```

Set control parameters.

```
fitCtrl <- trainControl(method = "repeatedcv",
                        number = 3,
                        repeats = 2,
                        search = "random")
```

Set testing grid.

```
glmnetGrid <- expand.grid(alpha=seq(0,1,by=0.1), lambda=seq(0,1,by=0.05))
```

Train model.

```
glmnet.res <- train(PSYCDX ~ .,
                   data=train,
                   method="glmnet",
                   trControl=fitCtrl,
                   tuneGrid=glmnetGrid,
                   metric="Accuracy")
```

```
## Warning in lognet(xd, is.sparse, ix, jx, y, weights, offset, alpha, nob, : one
## multinomial or binomial class has fewer than 8 observations; dangerous ground
```

```
## Warning in lognet(xd, is.sparse, ix, jx, y, weights, offset, alpha, nob, : one
## multinomial or binomial class has fewer than 8 observations; dangerous ground
```

```
## Warning in lognet(xd, is.sparse, ix, jx, y, weights, offset, alpha, nob, : one
## multinomial or binomial class has fewer than 8 observations; dangerous ground
```

```
## Warning in lognet(xd, is.sparse, ix, jx, y, weights, offset, alpha, nob, : one
## multinomial or binomial class has fewer than 8 observations; dangerous ground
```

```
## Warning in lognet(xd, is.sparse, ix, jx, y, weights, offset, alpha, nob, : one
## multinomial or binomial class has fewer than 8 observations; dangerous ground
```

```
## Warning in lognet(xd, is.sparse, ix, jx, y, weights, offset, alpha, nob, : one
## multinomial or binomial class has fewer than 8 observations; dangerous ground
```



```
## Warning in lognet(xd, is.sparse, ix, jx, y, weights, offset, alpha, nobs, : one
## multinomial or binomial class has fewer than 8 observations; dangerous ground

## Warning in lognet(xd, is.sparse, ix, jx, y, weights, offset, alpha, nobs, : one
## multinomial or binomial class has fewer than 8 observations; dangerous ground

## Warning in lognet(xd, is.sparse, ix, jx, y, weights, offset, alpha, nobs, : one
## multinomial or binomial class has fewer than 8 observations; dangerous ground

## Warning in lognet(xd, is.sparse, ix, jx, y, weights, offset, alpha, nobs, : one
## multinomial or binomial class has fewer than 8 observations; dangerous ground

## Warning in lognet(xd, is.sparse, ix, jx, y, weights, offset, alpha, nobs, : one
## multinomial or binomial class has fewer than 8 observations; dangerous ground
```

```
glmnet.res
```

```
## glmnet
##
## 146 samples
## 6 predictor
## 2 classes: 'non_psy', 'psy'
##
## No pre-processing
## Resampling: Cross-Validated (3 fold, repeated 2 times)
## Summary of sample sizes: 98, 97, 97, 98, 96, 98, ...
## Resampling results across tuning parameters:
##
##  alpha  lambda  Accuracy  Kappa
##  0.0    0.00   0.9247336 -0.01063950
##  0.0    0.05   0.9316071  0.00000000
##  0.0    0.10   0.9316071  0.00000000
##  0.0    0.15   0.9316071  0.00000000
##  0.0    0.20   0.9316071  0.00000000
##  0.0    0.25   0.9316071  0.00000000
##  0.0    0.30   0.9316071  0.00000000
##  0.0    0.35   0.9316071  0.00000000
##  0.0    0.40   0.9316071  0.00000000
##  0.0    0.45   0.9316071  0.00000000
##  0.0    0.50   0.9316071  0.00000000
##  0.0    0.55   0.9316071  0.00000000
##  0.0    0.60   0.9316071  0.00000000
##  0.0    0.65   0.9316071  0.00000000
##  0.0    0.70   0.9316071  0.00000000
##  0.0    0.75   0.9316071  0.00000000
##  0.0    0.80   0.9316071  0.00000000
##  0.0    0.85   0.9316071  0.00000000
##  0.0    0.90   0.9316071  0.00000000
##  0.0    0.95   0.9316071  0.00000000
##  0.0    1.00   0.9316071  0.00000000
##  0.1    0.00   0.8972392  0.00670904
```

##	0.1	0.05	0.9316071	0.00000000
##	0.1	0.10	0.9316071	0.00000000
##	0.1	0.15	0.9316071	0.00000000
##	0.1	0.20	0.9316071	0.00000000
##	0.1	0.25	0.9316071	0.00000000
##	0.1	0.30	0.9316071	0.00000000
##	0.1	0.35	0.9316071	0.00000000
##	0.1	0.40	0.9316071	0.00000000
##	0.1	0.45	0.9316071	0.00000000
##	0.1	0.50	0.9316071	0.00000000
##	0.1	0.55	0.9316071	0.00000000
##	0.1	0.60	0.9316071	0.00000000
##	0.1	0.65	0.9316071	0.00000000
##	0.1	0.70	0.9316071	0.00000000
##	0.1	0.75	0.9316071	0.00000000
##	0.1	0.80	0.9316071	0.00000000
##	0.1	0.85	0.9316071	0.00000000
##	0.1	0.90	0.9316071	0.00000000
##	0.1	0.95	0.9316071	0.00000000
##	0.1	1.00	0.9316071	0.00000000
##	0.2	0.00	0.8972392	0.00670904
##	0.2	0.05	0.9316071	0.00000000
##	0.2	0.10	0.9316071	0.00000000
##	0.2	0.15	0.9316071	0.00000000
##	0.2	0.20	0.9316071	0.00000000
##	0.2	0.25	0.9316071	0.00000000
##	0.2	0.30	0.9316071	0.00000000
##	0.2	0.35	0.9316071	0.00000000
##	0.2	0.40	0.9316071	0.00000000
##	0.2	0.45	0.9316071	0.00000000
##	0.2	0.50	0.9316071	0.00000000
##	0.2	0.55	0.9316071	0.00000000
##	0.2	0.60	0.9316071	0.00000000
##	0.2	0.65	0.9316071	0.00000000
##	0.2	0.70	0.9316071	0.00000000
##	0.2	0.75	0.9316071	0.00000000
##	0.2	0.80	0.9316071	0.00000000
##	0.2	0.85	0.9316071	0.00000000
##	0.2	0.90	0.9316071	0.00000000
##	0.2	0.95	0.9316071	0.00000000
##	0.2	1.00	0.9316071	0.00000000
##	0.3	0.00	0.8972392	0.00670904
##	0.3	0.05	0.9316071	0.00000000
##	0.3	0.10	0.9316071	0.00000000
##	0.3	0.15	0.9316071	0.00000000
##	0.3	0.20	0.9316071	0.00000000
##	0.3	0.25	0.9316071	0.00000000
##	0.3	0.30	0.9316071	0.00000000
##	0.3	0.35	0.9316071	0.00000000
##	0.3	0.40	0.9316071	0.00000000
##	0.3	0.45	0.9316071	0.00000000
##	0.3	0.50	0.9316071	0.00000000
##	0.3	0.55	0.9316071	0.00000000
##	0.3	0.60	0.9316071	0.00000000

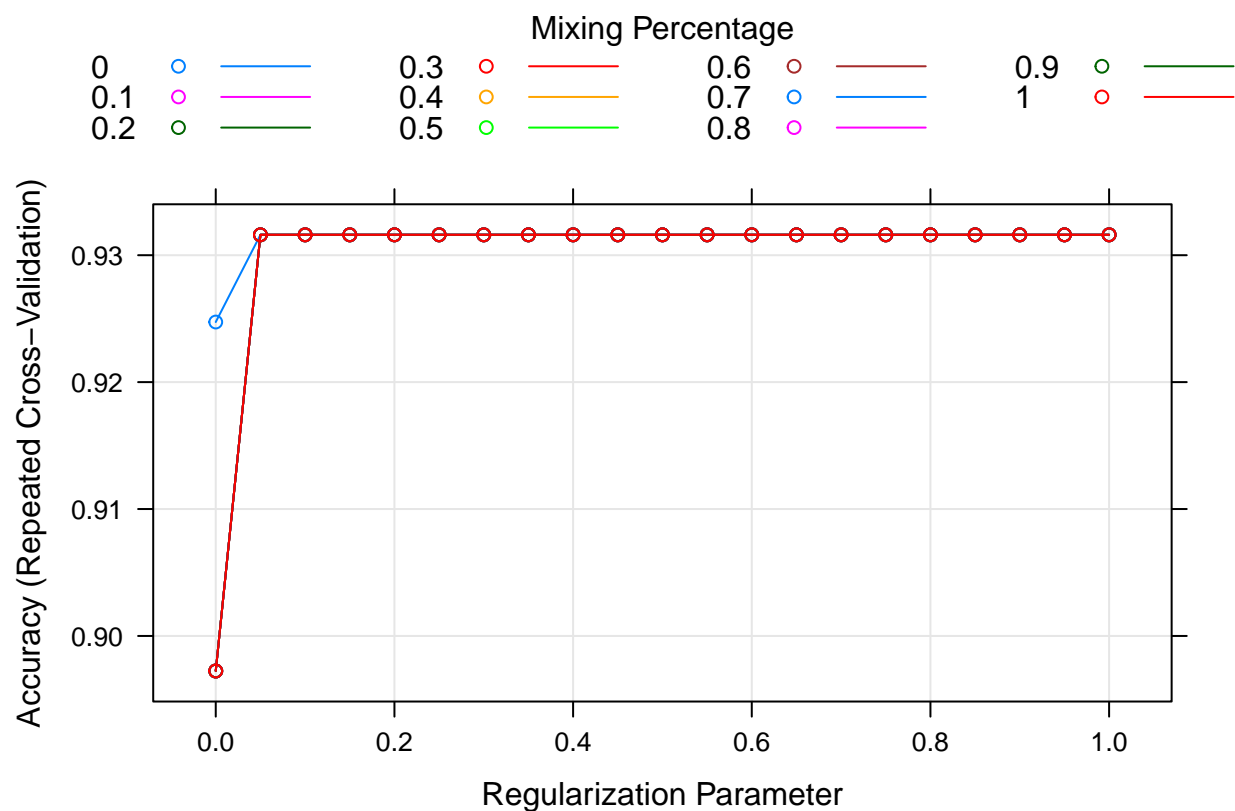
##	0.3	0.65	0.9316071	0.00000000
##	0.3	0.70	0.9316071	0.00000000
##	0.3	0.75	0.9316071	0.00000000
##	0.3	0.80	0.9316071	0.00000000
##	0.3	0.85	0.9316071	0.00000000
##	0.3	0.90	0.9316071	0.00000000
##	0.3	0.95	0.9316071	0.00000000
##	0.3	1.00	0.9316071	0.00000000
##	0.4	0.00	0.8972392	0.00670904
##	0.4	0.05	0.9316071	0.00000000
##	0.4	0.10	0.9316071	0.00000000
##	0.4	0.15	0.9316071	0.00000000
##	0.4	0.20	0.9316071	0.00000000
##	0.4	0.25	0.9316071	0.00000000
##	0.4	0.30	0.9316071	0.00000000
##	0.4	0.35	0.9316071	0.00000000
##	0.4	0.40	0.9316071	0.00000000
##	0.4	0.45	0.9316071	0.00000000
##	0.4	0.50	0.9316071	0.00000000
##	0.4	0.55	0.9316071	0.00000000
##	0.4	0.60	0.9316071	0.00000000
##	0.4	0.65	0.9316071	0.00000000
##	0.4	0.70	0.9316071	0.00000000
##	0.4	0.75	0.9316071	0.00000000
##	0.4	0.80	0.9316071	0.00000000
##	0.4	0.85	0.9316071	0.00000000
##	0.4	0.90	0.9316071	0.00000000
##	0.4	0.95	0.9316071	0.00000000
##	0.4	1.00	0.9316071	0.00000000
##	0.5	0.00	0.8972392	0.00670904
##	0.5	0.05	0.9316071	0.00000000
##	0.5	0.10	0.9316071	0.00000000
##	0.5	0.15	0.9316071	0.00000000
##	0.5	0.20	0.9316071	0.00000000
##	0.5	0.25	0.9316071	0.00000000
##	0.5	0.30	0.9316071	0.00000000
##	0.5	0.35	0.9316071	0.00000000
##	0.5	0.40	0.9316071	0.00000000
##	0.5	0.45	0.9316071	0.00000000
##	0.5	0.50	0.9316071	0.00000000
##	0.5	0.55	0.9316071	0.00000000
##	0.5	0.60	0.9316071	0.00000000
##	0.5	0.65	0.9316071	0.00000000
##	0.5	0.70	0.9316071	0.00000000
##	0.5	0.75	0.9316071	0.00000000
##	0.5	0.80	0.9316071	0.00000000
##	0.5	0.85	0.9316071	0.00000000
##	0.5	0.90	0.9316071	0.00000000
##	0.5	0.95	0.9316071	0.00000000
##	0.5	1.00	0.9316071	0.00000000
##	0.6	0.00	0.8972392	0.00670904
##	0.6	0.05	0.9316071	0.00000000
##	0.6	0.10	0.9316071	0.00000000
##	0.6	0.15	0.9316071	0.00000000

##	0.6	0.20	0.9316071	0.00000000
##	0.6	0.25	0.9316071	0.00000000
##	0.6	0.30	0.9316071	0.00000000
##	0.6	0.35	0.9316071	0.00000000
##	0.6	0.40	0.9316071	0.00000000
##	0.6	0.45	0.9316071	0.00000000
##	0.6	0.50	0.9316071	0.00000000
##	0.6	0.55	0.9316071	0.00000000
##	0.6	0.60	0.9316071	0.00000000
##	0.6	0.65	0.9316071	0.00000000
##	0.6	0.70	0.9316071	0.00000000
##	0.6	0.75	0.9316071	0.00000000
##	0.6	0.80	0.9316071	0.00000000
##	0.6	0.85	0.9316071	0.00000000
##	0.6	0.90	0.9316071	0.00000000
##	0.6	0.95	0.9316071	0.00000000
##	0.6	1.00	0.9316071	0.00000000
##	0.7	0.00	0.8972392	0.00670904
##	0.7	0.05	0.9316071	0.00000000
##	0.7	0.10	0.9316071	0.00000000
##	0.7	0.15	0.9316071	0.00000000
##	0.7	0.20	0.9316071	0.00000000
##	0.7	0.25	0.9316071	0.00000000
##	0.7	0.30	0.9316071	0.00000000
##	0.7	0.35	0.9316071	0.00000000
##	0.7	0.40	0.9316071	0.00000000
##	0.7	0.45	0.9316071	0.00000000
##	0.7	0.50	0.9316071	0.00000000
##	0.7	0.55	0.9316071	0.00000000
##	0.7	0.60	0.9316071	0.00000000
##	0.7	0.65	0.9316071	0.00000000
##	0.7	0.70	0.9316071	0.00000000
##	0.7	0.75	0.9316071	0.00000000
##	0.7	0.80	0.9316071	0.00000000
##	0.7	0.85	0.9316071	0.00000000
##	0.7	0.90	0.9316071	0.00000000
##	0.7	0.95	0.9316071	0.00000000
##	0.7	1.00	0.9316071	0.00000000
##	0.8	0.00	0.8972392	0.00670904
##	0.8	0.05	0.9316071	0.00000000
##	0.8	0.10	0.9316071	0.00000000
##	0.8	0.15	0.9316071	0.00000000
##	0.8	0.20	0.9316071	0.00000000
##	0.8	0.25	0.9316071	0.00000000
##	0.8	0.30	0.9316071	0.00000000
##	0.8	0.35	0.9316071	0.00000000
##	0.8	0.40	0.9316071	0.00000000
##	0.8	0.45	0.9316071	0.00000000
##	0.8	0.50	0.9316071	0.00000000
##	0.8	0.55	0.9316071	0.00000000
##	0.8	0.60	0.9316071	0.00000000
##	0.8	0.65	0.9316071	0.00000000
##	0.8	0.70	0.9316071	0.00000000
##	0.8	0.75	0.9316071	0.00000000

```

## 0.8 0.80 0.9316071 0.00000000
## 0.8 0.85 0.9316071 0.00000000
## 0.8 0.90 0.9316071 0.00000000
## 0.8 0.95 0.9316071 0.00000000
## 0.8 1.00 0.9316071 0.00000000
## 0.9 0.00 0.8972392 0.00670904
## 0.9 0.05 0.9316071 0.00000000
## 0.9 0.10 0.9316071 0.00000000
## 0.9 0.15 0.9316071 0.00000000
## 0.9 0.20 0.9316071 0.00000000
## 0.9 0.25 0.9316071 0.00000000
## 0.9 0.30 0.9316071 0.00000000
## 0.9 0.35 0.9316071 0.00000000
## 0.9 0.40 0.9316071 0.00000000
## 0.9 0.45 0.9316071 0.00000000
## 0.9 0.50 0.9316071 0.00000000
## 0.9 0.55 0.9316071 0.00000000
## 0.9 0.60 0.9316071 0.00000000
## 0.9 0.65 0.9316071 0.00000000
## 0.9 0.70 0.9316071 0.00000000
## 0.9 0.75 0.9316071 0.00000000
## 0.9 0.80 0.9316071 0.00000000
## 0.9 0.85 0.9316071 0.00000000
## 0.9 0.90 0.9316071 0.00000000
## 0.9 0.95 0.9316071 0.00000000
## 0.9 1.00 0.9316071 0.00000000
## 1.0 0.00 0.8972392 0.00670904
## 1.0 0.05 0.9316071 0.00000000
## 1.0 0.10 0.9316071 0.00000000
## 1.0 0.15 0.9316071 0.00000000
## 1.0 0.20 0.9316071 0.00000000
## 1.0 0.25 0.9316071 0.00000000
## 1.0 0.30 0.9316071 0.00000000
## 1.0 0.35 0.9316071 0.00000000
## 1.0 0.40 0.9316071 0.00000000
## 1.0 0.45 0.9316071 0.00000000
## 1.0 0.50 0.9316071 0.00000000
## 1.0 0.55 0.9316071 0.00000000
## 1.0 0.60 0.9316071 0.00000000
## 1.0 0.65 0.9316071 0.00000000
## 1.0 0.70 0.9316071 0.00000000
## 1.0 0.75 0.9316071 0.00000000
## 1.0 0.80 0.9316071 0.00000000
## 1.0 0.85 0.9316071 0.00000000
## 1.0 0.90 0.9316071 0.00000000
## 1.0 0.95 0.9316071 0.00000000
## 1.0 1.00 0.9316071 0.00000000
##
## Accuracy was used to select the optimal model using the largest value.
## The final values used for the model were alpha = 0 and lambda = 1.
plot(glmnet.res)

```



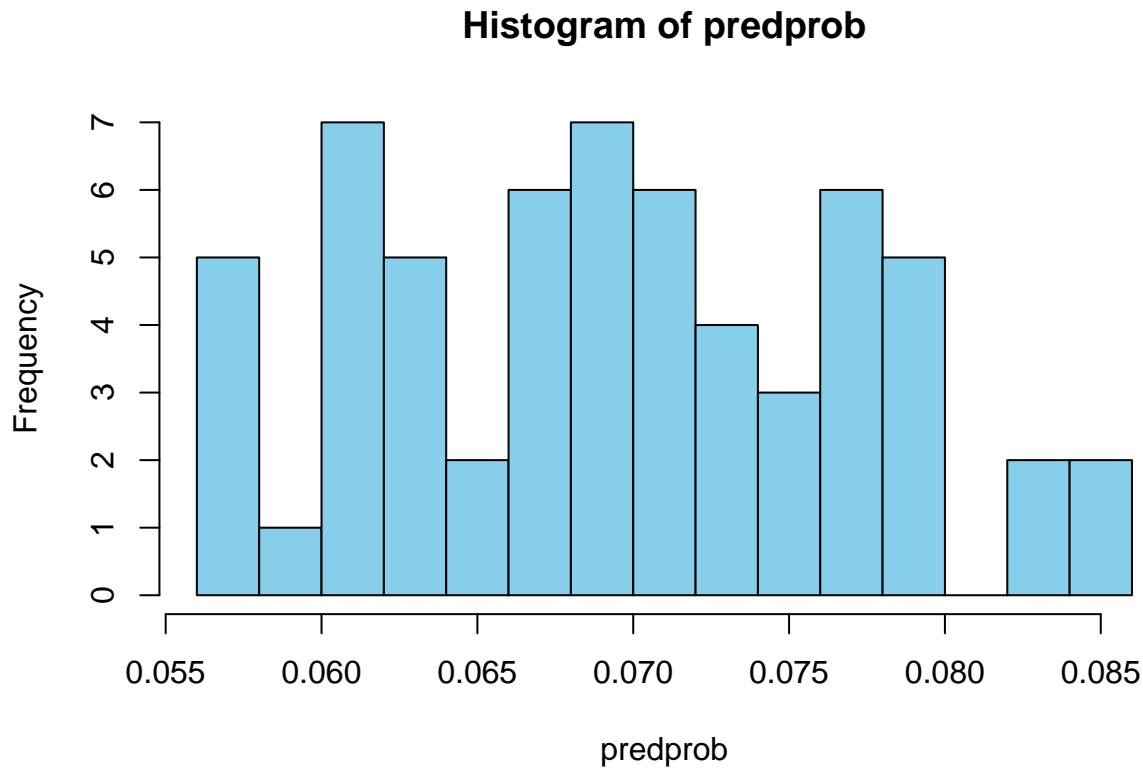
The optimal values are $\alpha = 0$ and $\lambda = 1$. Since α is 0, this elastic net is ridge regression not lasso regression.

Model performance.

```
predclass <- predict(glmnet.res, test)
table(predclass, test$PSYCDX[complete.cases(test)])
```

```
##
## predclass non_psy psy
## non_psy      58   3
## psy          0   0
```

```
predprob <- predict(glmnet.res, test, type="prob")[,"psy"]
hist(predprob, col="skyblue", breaks=20)
```



```
roc(test$PSYCDX[complete.cases(test)] ~ predprob)
```

```
## Setting levels: control = non_psy, case = psy
```

```
## Setting direction: controls < cases
```

```
##
```

```
## Call:
```

```
## roc.formula(formula = test$PSYCDX[complete.cases(test)] ~ predprob)
```

```
##
```

```
## Data: predprob in 58 controls (test$PSYCDX[complete.cases(test)] non_psy) < 3 cases (test$PSYCDX[complete.cases(test)] psy)
```

```
## Area under the curve: 1
```

I got an area under the curve of 1...that really makes me feel like something went wrong because I highly doubt I have a perfect model. Let's keep moving forward.

Here we find the beta coefficients.

```
fit.elasticnet <- glmnet(as.matrix(train[,-1]), as.numeric(train[,1]), family="binomial", alpha=0, lambda=1e-5)
fit.elasticnet$beta
```

```
## 6 x 1 sparse Matrix of class "dgCMatrix"
```

```
##              s0
```

```
## SEX          0.060431601
```

```
## SES          0.028883468
```

```
## MOODDX       -0.030461034
```

```
## SCL_PAR      0.009019624
```

```
## SISDB_SUB    0.043673927
```

```
## AGE         -0.042595333
```

Since my results suggested a ridge regression, none of my coefficients are actually shrunk to zero and that's expected. So, all of my predictors are included in the model.