hw1

Amanda Howarth

2/26/2020

```
library(tidyverse)
```

```
## ── Attaching packages
─────────────────────────────────────────────────────── tidyverse 1.2.1 ──
```

```
## ✓ ggplot2 3.2.1     ✓ purrr   0.3.2
## ✓ tibble  2.1.3     ✓ dplyr   0.8.3
## ✓ tidyr   1.0.0     ✓ stringr 1.4.0
## ✓ readr   1.3.1     ✓ forcats 0.4.0
```

```
## ── Conflicts ──────────────────────────────────────────────────────────
tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
```

```
library(readxl)
library(stargazer)
```

```
##
## Please cite as:
```

```
##  Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary
Statistics Tables.
```

```
##  R package version 5.2.2. https://CRAN.R-project.org/package=stargazer
```

```
library(leaps)
library(FNN)
library(ModelMetrics)
```

```
##
## Attaching package: 'ModelMetrics'
```

```
## The following object is masked from 'package:base':
##
##     kappa
```

```
library(caret)
```

```
## Loading required package: lattice
```

```
##
## Attaching package: 'caret'
```

```
## The following objects are masked from 'package:ModelMetrics':
##
##     confusionMatrix, precision, recall, sensitivity, specificity

## The following object is masked from 'package:purrr':
##
##     lift

library(boot)

##
## Attaching package: 'boot'

## The following object is masked from 'package:lattice':
##
##     melanoma

library(Rcpp)
library(microbenchmark)
library(ISLR)
library(glmnet)

## Loading required package: Matrix

##
## Attaching package: 'Matrix'

## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack

## Loaded glmnet 3.0-2

library(corrplot)

## corrplot 0.84 loaded

library(plotmo)

## Loading required package: Formula

## Loading required package: plotrix

## Loading required package: TeachingDemos

test =
  read_excel(path = "./data/solubility_test.xlsx", sheet = 1) %>%
  janitor::clean_names() %>%
  na.omit()

train =
  read_excel(path = "./data/solubility_train.xlsx", sheet = 1) %>%
```

```
  janitor::clean_names() %>%
  na.omit()
```

#QUSETION 1A 1. (a) Fit a linear model using least squares on the training data and calculate the mean square error using the test data.

```
ctrl1 <- trainControl(method = "repeatedcv", number = 10, repeats = 5)

set.seed(2)
lmfit <- train(solubility~.,
                data = train,
                method = "lm",
                trControl = ctrl1)

## Warning in predict.lm(modelFit, newdata): prediction from a rank-deficient
## fit may be misleading

## Warning in predict.lm(modelFit, newdata): prediction from a rank-deficient
## fit may be misleading

## Warning in predict.lm(modelFit, newdata): prediction from a rank-deficient
## fit may be misleading

## Warning in predict.lm(modelFit, newdata): prediction from a rank-deficient
## fit may be misleading

## Warning in predict.lm(modelFit, newdata): prediction from a rank-deficient
## fit may be misleading

## Warning in predict.lm(modelFit, newdata): prediction from a rank-deficient
## fit may be misleading

## Warning in predict.lm(modelFit, newdata): prediction from a rank-deficient
## fit may be misleading

## Warning in predict.lm(modelFit, newdata): prediction from a rank-deficient
## fit may be misleading

## Warning in predict.lm(modelFit, newdata): prediction from a rank-deficient
## fit may be misleading

lmfit

## Linear Regression
##
## 951 samples
## 228 predictors
##
## No pre-processing
## Resampling: Cross-Validated (10 fold, repeated 5 times)
```

```
## Summary of sample sizes: 855, 856, 855, 855, 856, 856, ...
## Resampling results:
##
##   RMSE       Rsquared   MAE
##   0.7093576  0.8814378  0.530454
##
## Tuning parameter 'intercept' was held constant at a value of TRUE
```

```
#test error
pred_lm <- predict(lmfit, test)
mse(test$solubility, pred_lm)
```

```
## [1] 0.5558898
```

Using the test data, we find that the mean square error (MSE) is 0.5558898.

#QUSETION 1B 1b. Fit a ridge regression model on the training data, with λ chosen by cross-validation. Report the test error.

```
x <- model.matrix(solubility~.,train)[,-1]
y <- train$solubility

ctrl1 <- trainControl(method = "repeatedcv", number = 10, repeats = 5)

set.seed(2)
ridge.fit <- train(x, y,
                   method = "glmnet",
                   tuneGrid = expand.grid(alpha = 0,
                                          lambda = exp(seq(-50, 50,
length=100))),
                   preProc = c("center", "scale"),
                   trControl = ctrl1)
```

```
## Warning in nominalTrainWorkflow(x = x, y = y, wts = weights, info =
## trainInfo, : There were missing values in resampled performance measures.
```

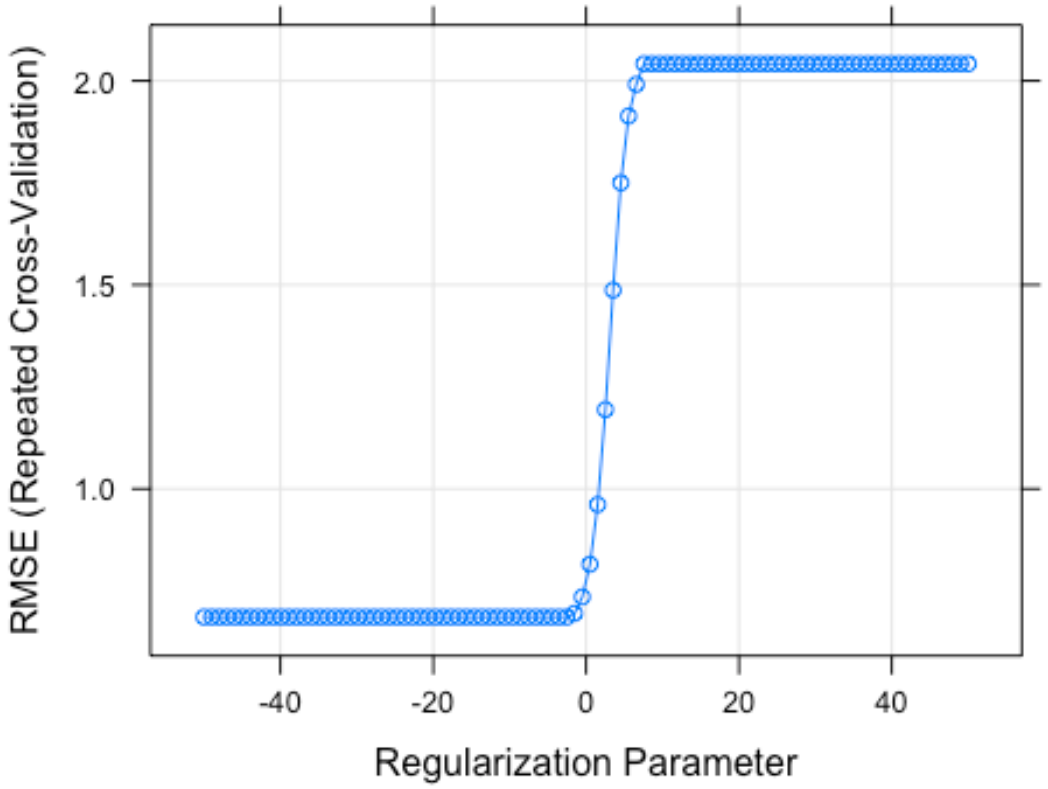
```
ridge.fit
```

```
## glmnet
##
## 951 samples
## 228 predictors
##
## Pre-processing: centered (228), scaled (228)
## Resampling: Cross-Validated (10 fold, repeated 5 times)
## Summary of sample sizes: 855, 856, 855, 855, 856, 856, ...
## Resampling results across tuning parameters:
##
##   lambda        RMSE       Rsquared   MAE
##   1.928750e-22  0.6856755  0.8880644  0.5213242
##   5.296112e-22  0.6856755  0.8880644  0.5213242
```

```
##   1.454248e-21  0.6856755  0.8880644  0.5213242
##   3.993188e-21  0.6856755  0.8880644  0.5213242
##   1.096481e-20  0.6856755  0.8880644  0.5213242
##   3.010803e-20  0.6856755  0.8880644  0.5213242
##   8.267300e-20  0.6856755  0.8880644  0.5213242
##   2.270100e-19  0.6856755  0.8880644  0.5213242
##   6.233418e-19  0.6856755  0.8880644  0.5213242
##   1.711621e-18  0.6856755  0.8880644  0.5213242
##   4.699903e-18  0.6856755  0.8880644  0.5213242
##   1.290536e-17  0.6856755  0.8880644  0.5213242
##   3.543655e-17  0.6856755  0.8880644  0.5213242
##   9.730446e-17  0.6856755  0.8880644  0.5213242
##   2.671862e-16  0.6856755  0.8880644  0.5213242
##   7.336608e-16  0.6856755  0.8880644  0.5213242
##   2.014543e-15  0.6856755  0.8880644  0.5213242
##   5.531691e-15  0.6856755  0.8880644  0.5213242
##   1.518935e-14  0.6856755  0.8880644  0.5213242
##   4.170811e-14  0.6856755  0.8880644  0.5213242
##   1.145254e-13  0.6856755  0.8880644  0.5213242
##   3.144728e-13  0.6856755  0.8880644  0.5213242
##   8.635041e-13  0.6856755  0.8880644  0.5213242
##   2.371077e-12  0.6856755  0.8880644  0.5213242
##   6.510689e-12  0.6856755  0.8880644  0.5213242
##   1.787756e-11  0.6856755  0.8880644  0.5213242
##   4.908961e-11  0.6856755  0.8880644  0.5213242
##   1.347941e-10  0.6856755  0.8880644  0.5213242
##   3.701282e-10  0.6856755  0.8880644  0.5213242
##   1.016327e-09  0.6856755  0.8880644  0.5213242
##   2.790710e-09  0.6856755  0.8880644  0.5213242
##   7.662951e-09  0.6856755  0.8880644  0.5213242
##   2.104153e-08  0.6856755  0.8880644  0.5213242
##   5.777749e-08  0.6856755  0.8880644  0.5213242
##   1.586499e-07  0.6856755  0.8880644  0.5213242
##   4.356335e-07  0.6856755  0.8880644  0.5213242
##   1.196196e-06  0.6856755  0.8880644  0.5213242
##   3.284610e-06  0.6856755  0.8880644  0.5213242
##   9.019140e-06  0.6856755  0.8880644  0.5213242
##   2.476546e-05  0.6856755  0.8880644  0.5213242
##   6.800294e-05  0.6856755  0.8880644  0.5213242
##   1.867278e-04  0.6856755  0.8880644  0.5213242
##   5.127318e-04  0.6856755  0.8880644  0.5213242
##   1.407899e-03  0.6856755  0.8880644  0.5213242
##   3.865920e-03  0.6856755  0.8880644  0.5213242
##   1.061535e-02  0.6856755  0.8880644  0.5213242
##   2.914845e-02  0.6856755  0.8880644  0.5213242
##   8.003810e-02  0.6856755  0.8880644  0.5213242
##   2.197749e-01  0.6947539  0.8854305  0.5284826
##   6.034751e-01  0.7349791  0.8739060  0.5622516
##   1.657069e+00  0.8150841  0.8524399  0.6283805
##   4.550110e+00  0.9613507  0.8166167  0.7415463
```

```
##   1.249405e+01  1.1942179  0.7651233  0.9165596
##   3.430714e+01  1.4870528  0.7076171  1.1457491
##   9.420324e+01  1.7500053  0.6597887  1.3573863
##   2.586706e+02  1.9138531  0.6306032  1.4888828
##   7.102781e+02  1.9914139  0.6168452  1.5503534
##   1.950337e+03  2.0419201        NaN  1.5903934
##   5.355389e+03  2.0419201        NaN  1.5903934
##   1.470525e+04  2.0419201        NaN  1.5903934
##   4.037882e+04  2.0419201        NaN  1.5903934
##   1.108753e+05  2.0419201        NaN  1.5903934
##   3.044501e+05  2.0419201        NaN  1.5903934
##   8.359831e+05  2.0419201        NaN  1.5903934
##   2.295508e+06  2.0419201        NaN  1.5903934
##   6.303185e+06  2.0419201        NaN  1.5903934
##   1.730778e+07  2.0419201        NaN  1.5903934
##   4.752506e+07  2.0419201        NaN  1.5903934
##   1.304980e+08  2.0419201        NaN  1.5903934
##   3.583317e+08  2.0419201        NaN  1.5903934
##   9.839353e+08  2.0419201        NaN  1.5903934
##   2.701767e+09  2.0419201        NaN  1.5903934
##   7.418723e+09  2.0419201        NaN  1.5903934
##   2.037091e+10  2.0419201        NaN  1.5903934
##   5.593604e+10  2.0419201        NaN  1.5903934
##   1.535936e+11  2.0419201        NaN  1.5903934
##   4.217492e+11  2.0419201        NaN  1.5903934
##   1.158072e+12  2.0419201        NaN  1.5903934
##   3.179925e+12  2.0419201        NaN  1.5903934
##   8.731688e+12  2.0419201        NaN  1.5903934
##   2.397615e+13  2.0419201        NaN  1.5903934
##   6.583560e+13  2.0419201        NaN  1.5903934
##   1.807765e+14  2.0419201        NaN  1.5903934
##   4.963904e+14  2.0419201        NaN  1.5903934
##   1.363028e+15  2.0419201        NaN  1.5903934
##   3.742708e+15  2.0419201        NaN  1.5903934
##   1.027702e+16  2.0419201        NaN  1.5903934
##   2.821945e+16  2.0419201        NaN  1.5903934
##   7.748718e+16  2.0419201        NaN  1.5903934
##   2.127704e+17  2.0419201        NaN  1.5903934
##   5.842416e+17  2.0419201        NaN  1.5903934
##   1.604256e+18  2.0419201        NaN  1.5903934
##   4.405092e+18  2.0419201        NaN  1.5903934
##   1.209585e+19  2.0419201        NaN  1.5903934
##   3.321373e+19  2.0419201        NaN  1.5903934
##   9.120086e+19  2.0419201        NaN  1.5903934
##   2.504265e+20  2.0419201        NaN  1.5903934
##   6.876406e+20  2.0419201        NaN  1.5903934
##   1.888177e+21  2.0419201        NaN  1.5903934
##   5.184706e+21  2.0419201        NaN  1.5903934
##
## Tuning parameter 'alpha' was held constant at a value of 0
```

```
## RMSE was used to select the optimal model using the smallest value.
## The final values used for the model were alpha = 0 and lambda = 0.0800381.

plot(ridge.fit, xTrans = function(x) log(x))
```




```
#lambda value
ridge.fit$bestTune

##    alpha    lambda
## 48     0 0.0800381

#model coefficients
coef(ridge.fit$finalModel,ridge.fit$bestTune$lambda)

## 229 x 1 sparse Matrix of class "dgCMatrix"
##                                    1
## (Intercept)        -2.7185699264
## fp001               0.0182944967
## fp002               0.0647233005
## fp003              -0.0370236482
## fp004              -0.0855575442
## fp005              -0.0133319156
## fp006              -0.0484578041
## fp007               0.0171343338
```

```
## fp008               0.0154028137
## fp009              -0.0109157478
## fp010               0.0146349259
## fp011               0.0430879555
## fp012              -0.0252151727
## fp013              -0.0319703472
## fp014               0.0027293475
## fp015              -0.0095942577
## fp016              -0.0415782969
## fp017              -0.0663470313
## fp018              -0.0528802638
## fp019               0.0062105219
## fp020               0.0410118164
## fp021               0.0032185171
## fp022               0.0465650184
## fp023              -0.0536614672
## fp024              -0.0402200374
## fp025               0.0048713425
## fp026               0.0454909264
## fp027               0.0662206315
## fp028               0.0225095264
## fp029              -0.0168278955
## fp030              -0.0480125247
## fp031               0.0557571797
## fp032              -0.0441088234
## fp033               0.0790809785
## fp034              -0.0480994528
## fp035              -0.0353837951
## fp036              -0.0021718104
## fp037               0.0695178965
## fp038               0.0389560992
## fp039              -0.1101428983
## fp040               0.1181431694
## fp041              -0.0074271387
## fp042               0.0099370784
## fp043               0.0161763635
## fp044              -0.0749506941
## fp045               0.0273431814
## fp046               0.0486001510
## fp047              -0.0226012570
## fp048               0.0116649452
## fp049               0.0996149434
## fp050              -0.0304484965
## fp051              -0.0089965432
## fp052              -0.0285636622
## fp053               0.0944727800
## fp054              -0.0234136135
## fp055              -0.0387318060
## fp056              -0.0165622829
## fp057              -0.0377558140
```

```
## fp058               0.0149290668
## fp059              -0.0724769694
## fp060               0.0514859154
## fp061              -0.0366074133
## fp062               0.0392198459
## fp063               0.1105865213
## fp064               0.0741610549
## fp065              -0.0788714728
## fp066               0.0626316555
## fp067              -0.0464873861
## fp068               0.0353466639
## fp069               0.0718691507
## fp070              -0.0933489007
## fp071               0.0553084164
## fp072               0.1394235371
## fp073              -0.0467871980
## fp074               0.0609520559
## fp075               0.1117110297
## fp076               0.0161207230
## fp077               0.0469486084
## fp078              -0.0166586103
## fp079               0.0716951429
## fp080               0.0511455405
## fp081              -0.0772665525
## fp082               0.0467061992
## fp083              -0.1172124196
## fp084               0.0877208784
## fp085              -0.1284733773
## fp086              -0.0306711400
## fp087               0.0226928808
## fp088               0.0631432127
## fp089              -0.0628035304
## fp090              -0.0174046315
## fp091               0.0204204398
## fp092               0.0113111852
## fp093               0.0515084724
## fp094              -0.0453726216
## fp095               0.0147884043
## fp096               0.0210278964
## fp097               0.0480291117
## fp098              -0.0241339746
## fp099               0.0628406059
## fp100              -0.0315018128
## fp101               0.0413788246
## fp102               0.0527641877
## fp103              -0.0553405300
## fp104              -0.0522248112
## fp105              -0.0329472116
## fp106               0.0030951658
## fp107              -0.0057300479
```

```
## fp108               0.0190123907
## fp109               0.0960092909
## fp110               0.0125294737
## fp111              -0.1477342576
## fp112              -0.0427900086
## fp113               0.0347223479
## fp114               0.0435445509
## fp115               0.0471929563
## fp116               0.0609334904
## fp117              -0.0398176785
## fp118              -0.0534524608
## fp119               0.0800253061
## fp120              -0.0234887668
## fp121              -0.0353530559
## fp122               0.0716608687
## fp123              -0.0141193357
## fp124               0.0848870880
## fp125               0.0322061122
## fp126              -0.1096426378
## fp127              -0.1048117874
## fp128              -0.0793231369
## fp129               0.0190499929
## fp130              -0.0776383348
## fp131               0.0917771426
## fp132              -0.0157863528
## fp133              -0.0443952035
## fp134              -0.0367974890
## fp135               0.0584291002
## fp136               0.0329490439
## fp137               0.0262628186
## fp138               0.0562301365
## fp139               0.0115837076
## fp140               0.0268059723
## fp141              -0.0475812786
## fp142               0.1098076669
## fp143               0.0862783656
## fp144               0.0091386512
## fp145              -0.0620902591
## fp146              -0.0276692949
## fp147               0.0736553748
## fp148               0.0018266597
## fp149              -0.0069699804
## fp150               0.0365715706
## fp151               0.0276135807
## fp152              -0.0050110810
## fp153              -0.0275939273
## fp154              -0.0672354583
## fp155               0.0330976723
## fp156              -0.0590177068
## fp157               0.0078664067
```

```
## fp158              -0.0092541295
## fp159               0.0665157726
## fp160              -0.0337212595
## fp161              -0.0156930702
## fp162               0.0315361418
## fp163               0.1125009598
## fp164               0.1218942620
## fp165              -0.0161739149
## fp166               0.0458110229
## fp167              -0.0657933778
## fp168              -0.0611707124
## fp169              -0.0646652814
## fp170               0.0231572946
## fp171               0.0896673806
## fp172              -0.1158353634
## fp173               0.1134597926
## fp174              -0.0405712164
## fp175              -0.0149263283
## fp176               0.1086053804
## fp177              -0.0035802379
## fp178              -0.0037931209
## fp179               0.0236633704
## fp180              -0.0354078857
## fp181               0.0291062953
## fp182              -0.0138883720
## fp183              -0.0039064624
## fp184               0.0762092432
## fp185              -0.0198161545
## fp186              -0.0625490812
## fp187               0.0377749992
## fp188               0.0616741200
## fp189               0.0131773729
## fp190               0.0522285022
## fp191               0.0184264239
## fp192               0.0177576034
## fp193              -0.0178358968
## fp194               0.0077458076
## fp195              -0.0099503282
## fp196               0.0192888160
## fp197              -0.0001049384
## fp198               0.0483601559
## fp199              -0.0025640841
## fp200              -0.0189922792
## fp201              -0.0620408524
## fp202               0.1079685339
## fp203               0.0186333392
## fp204              -0.0246045110
## fp205              -0.0306137989
## fp206              -0.0229748991
## fp207              -0.0092674578
```

```
## fp208                0.0018060022
## mol_weight          -0.5325668610
## num_atoms           -0.1853715091
## num_non_h_atoms     -0.2943620395
## num_bonds           -0.1707277488
## num_non_h_bonds     -0.2920202275
## num_mult_bonds      -0.1635462481
## num_rot_bonds       -0.1398561385
## num_dbl_bonds       -0.0119157747
## num_aromatic_bonds  -0.1296184809
## num_hydrogen         0.0997591041
## num_carbon          -0.2593023932
## num_nitrogen         0.1081757880
## num_oxygen           0.2287309627
## num_sulfer          -0.0766799439
## num_chlorine        -0.1152109404
## num_halogen         -0.0958922596
## num_rings           -0.1426644764
## hydrophilic_factor   0.1406610534
## surface_area1        0.3631751516
## surface_area2        0.2284711220
```

```
#test error
pred_ridge <- predict(ridge.fit, test)
mse(test$solubility, pred_ridge)
```

```
## [1] 0.5134603
```

We found the test erorr to be 0.5134603 with a lambda value of 0.0800381.

##Question 1C 1c. Fit a lasso model on the training data, with λ chosen by cross-validation. Report the test error, along with the number of non-zero coefficient estimates.

```
set.seed(2)
lasso.fit <- train(x, y,
                   method = "glmnet",
                   tuneGrid = expand.grid(alpha = 1,
                                          lambda = exp(seq(-10, 10,
length=100))),
                   preProc = c("center", "scale"),
                   trControl = ctrl1)
```

```
## Warning in nominalTrainWorkflow(x = x, y = y, wts = weights, info =
## trainInfo, : There were missing values in resampled performance measures.
```

```
lasso.fit
```

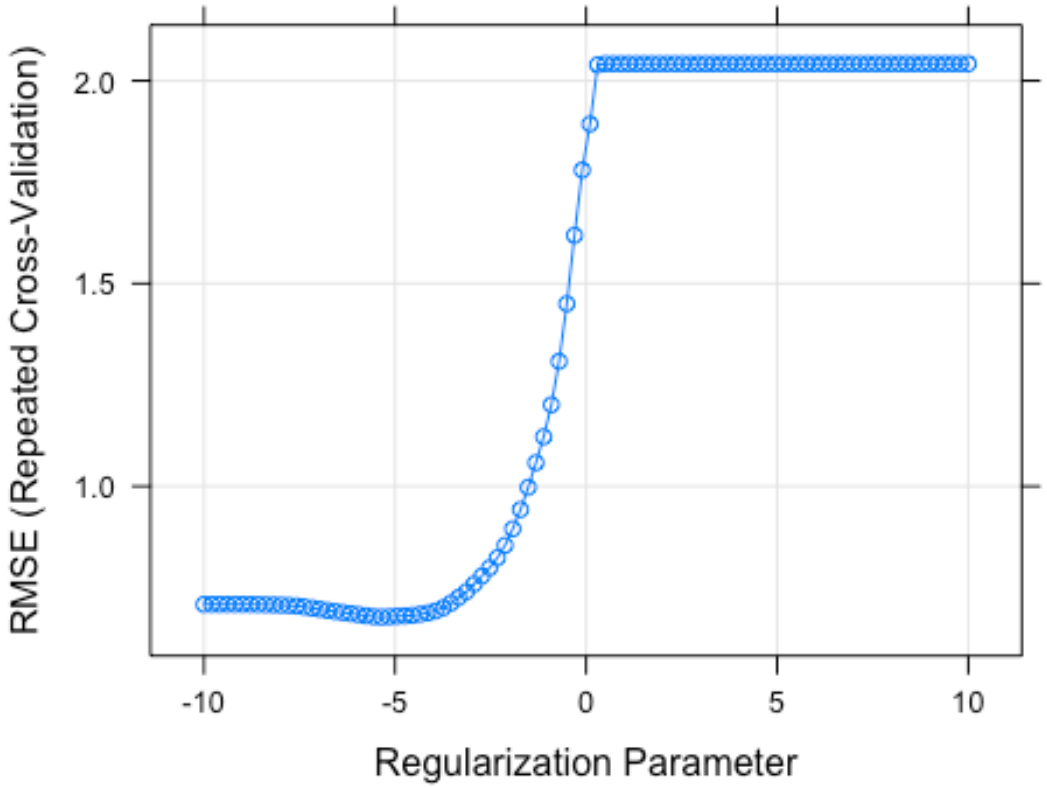
```
## glmnet
##
## 951 samples
## 228 predictors
```

```
##
## Pre-processing: centered (228), scaled (228)
## Resampling: Cross-Validated (10 fold, repeated 5 times)
## Summary of sample sizes: 855, 856, 855, 855, 856, 856, ...
## Resampling results across tuning parameters:
##
##   lambda        RMSE       Rsquared   MAE
##   4.539993e-05  0.7087425  0.8814440  0.5371863
##   5.556374e-05  0.7087425  0.8814440  0.5371863
##   6.800294e-05  0.7087425  0.8814440  0.5371863
##   8.322695e-05  0.7087425  0.8814440  0.5371863
##   1.018592e-04  0.7087425  0.8814440  0.5371863
##   1.246627e-04  0.7087425  0.8814440  0.5371863
##   1.525713e-04  0.7086700  0.8814554  0.5371318
##   1.867278e-04  0.7083228  0.8815424  0.5368644
##   2.285311e-04  0.7078476  0.8816599  0.5365257
##   2.796929e-04  0.7074314  0.8817660  0.5362170
##   3.423086e-04  0.7068793  0.8819177  0.5358245
##   4.189421e-04  0.7061306  0.8821258  0.5352521
##   5.127318e-04  0.7047502  0.8825313  0.5342299
##   6.275185e-04  0.7025083  0.8831910  0.5327641
##   7.680028e-04  0.6997113  0.8840165  0.5310837
##   9.399377e-04  0.6968110  0.8848715  0.5292456
##   1.150364e-03  0.6937842  0.8857586  0.5272310
##   1.407899e-03  0.6910188  0.8865502  0.5256187
##   1.723090e-03  0.6886085  0.8872217  0.5241559
##   2.108842e-03  0.6860677  0.8879704  0.5224912
##   2.580955e-03  0.6835242  0.8887228  0.5211648
##   3.158760e-03  0.6808284  0.8895238  0.5195994
##   3.865920e-03  0.6783560  0.8902446  0.5180268
##   4.731394e-03  0.6774968  0.8904842  0.5172918
##   5.790624e-03  0.6782049  0.8902294  0.5178885
##   7.086987e-03  0.6795133  0.8897894  0.5190690
##   8.673571e-03  0.6804230  0.8895115  0.5199633
##   1.061535e-02  0.6817907  0.8891464  0.5212810
##   1.299183e-02  0.6844911  0.8883619  0.5237590
##   1.590035e-02  0.6876928  0.8874794  0.5263750
##   1.946001e-02  0.6924562  0.8861694  0.5298823
##   2.381657e-02  0.7003631  0.8838840  0.5355245
##   2.914845e-02  0.7119032  0.8804184  0.5439375
##   3.567399e-02  0.7257980  0.8762012  0.5541241
##   4.366043e-02  0.7401892  0.8719199  0.5649201
##   5.343481e-02  0.7588173  0.8662455  0.5794530
##   6.539740e-02  0.7785811  0.8604880  0.5945621
##   8.003810e-02  0.7989045  0.8550632  0.6089010
##   9.795645e-02  0.8234826  0.8485763  0.6262212
##   1.198862e-01  0.8544675  0.8402498  0.6492219
##   1.467255e-01  0.8950644  0.8286673  0.6806922
##   1.795733e-01  0.9425401  0.8150378  0.7188665
##   2.197749e-01  0.9973932  0.7994156  0.7627176
```

```
##   2.689765e-01  1.0580632  0.7835697  0.8110282
##   3.291930e-01  1.1220567  0.7715695  0.8602303
##   4.028903e-01  1.2009410  0.7588376  0.9202956
##   4.930865e-01  1.3080810  0.7342346  1.0025399
##   6.034751e-01  1.4500299  0.6816862  1.1124295
##   7.385767e-01  1.6188646  0.5824416  1.2434223
##   9.039239e-01  1.7797663  0.4467333  1.3676461
##   1.106288e+00  1.8932043  0.4390840  1.4625919
##   1.353955e+00  2.0393003  0.3550235  1.5881301
##   1.657069e+00  2.0419201        NaN  1.5903934
##   2.028042e+00  2.0419201        NaN  1.5903934
##   2.482065e+00  2.0419201        NaN  1.5903934
##   3.037732e+00  2.0419201        NaN  1.5903934
##   3.717797e+00  2.0419201        NaN  1.5903934
##   4.550110e+00  2.0419201        NaN  1.5903934
##   5.568756e+00  2.0419201        NaN  1.5903934
##   6.815449e+00  2.0419201        NaN  1.5903934
##   8.341242e+00  2.0419201        NaN  1.5903934
##   1.020862e+01  2.0419201        NaN  1.5903934
##   1.249405e+01  2.0419201        NaN  1.5903934
##   1.529113e+01  2.0419201        NaN  1.5903934
##   1.871439e+01  2.0419201        NaN  1.5903934
##   2.290404e+01  2.0419201        NaN  1.5903934
##   2.803162e+01  2.0419201        NaN  1.5903934
##   3.430714e+01  2.0419201        NaN  1.5903934
##   4.198757e+01  2.0419201        NaN  1.5903934
##   5.138745e+01  2.0419201        NaN  1.5903934
##   6.289170e+01  2.0419201        NaN  1.5903934
##   7.697143e+01  2.0419201        NaN  1.5903934
##   9.420324e+01  2.0419201        NaN  1.5903934
##   1.152928e+02  2.0419201        NaN  1.5903934
##   1.411037e+02  2.0419201        NaN  1.5903934
##   1.726929e+02  2.0419201        NaN  1.5903934
##   2.113542e+02  2.0419201        NaN  1.5903934
##   2.586706e+02  2.0419201        NaN  1.5903934
##   3.165799e+02  2.0419201        NaN  1.5903934
##   3.874535e+02  2.0419201        NaN  1.5903934
##   4.741938e+02  2.0419201        NaN  1.5903934
##   5.803529e+02  2.0419201        NaN  1.5903934
##   7.102781e+02  2.0419201        NaN  1.5903934
##   8.692900e+02  2.0419201        NaN  1.5903934
##   1.063900e+03  2.0419201        NaN  1.5903934
##   1.302079e+03  2.0419201        NaN  1.5903934
##   1.593578e+03  2.0419201        NaN  1.5903934
##   1.950337e+03  2.0419201        NaN  1.5903934
##   2.386965e+03  2.0419201        NaN  1.5903934
##   2.921341e+03  2.0419201        NaN  1.5903934
##   3.575349e+03  2.0419201        NaN  1.5903934
##   4.375773e+03  2.0419201        NaN  1.5903934
##   5.355389e+03  2.0419201        NaN  1.5903934
```

```
##   6.554314e+03  2.0419201        NaN  1.5903934
##   8.021647e+03  2.0419201        NaN  1.5903934
##   9.817475e+03  2.0419201        NaN  1.5903934
##   1.201534e+04  2.0419201        NaN  1.5903934
##   1.470525e+04  2.0419201        NaN  1.5903934
##   1.799735e+04  2.0419201        NaN  1.5903934
##   2.202647e+04  2.0419201        NaN  1.5903934
##
## Tuning parameter 'alpha' was held constant at a value of 1
## RMSE was used to select the optimal model using the smallest value.
## The final values used for the model were alpha = 1 and lambda
##  = 0.004731394.
```

```
plot(lasso.fit, xTrans = function(x) log(x))
```




```
#lamda value
lasso.fit$bestTune
```

```
##    alpha      lambda
## 24     1 0.004731394
```

```
#model coefficients
coef(lasso.fit$finalModel,lasso.fit$bestTune$lambda)
```

```
## 229 x 1 sparse Matrix of class "dgCMatrix"
##                                 1
## (Intercept)        -2.718570e+00
## fp001               .
## fp002               1.197528e-01
## fp003              -2.082747e-02
## fp004              -1.141929e-01
## fp005               .
## fp006              -3.212338e-02
## fp007               .
## fp008               .
## fp009               .
## fp010               .
## fp011               .
## fp012              -1.835282e-02
## fp013              -2.313275e-02
## fp014               .
## fp015              -3.135592e-02
## fp016              -2.453394e-02
## fp017              -4.836466e-02
## fp018              -3.005309e-02
## fp019               .
## fp020               3.002163e-02
## fp021               .
## fp022               .
## fp023              -4.855186e-02
## fp024              -3.165895e-02
## fp025               .
## fp026               7.450465e-02
## fp027               8.879623e-02
## fp028               .
## fp029               .
## fp030              -4.614950e-02
## fp031               3.352798e-02
## fp032               .
## fp033               2.815407e-02
## fp034              -2.463474e-03
## fp035              -3.333394e-02
## fp036               .
## fp037               5.407199e-02
## fp038               1.729235e-02
## fp039              -1.119137e-01
## fp040               1.085222e-01
## fp041               .
## fp042               .
## fp043               1.411075e-02
## fp044              -7.095811e-02
## fp045               1.971281e-02
## fp046               .
## fp047               .
```

```
## fp048               .
## fp049               9.640206e-02
## fp050              -4.921156e-02
## fp051               .
## fp052               .
## fp053               7.013322e-02
## fp054              -2.082958e-02
## fp055              -3.020524e-02
## fp056               .
## fp057              -2.941848e-02
## fp058               .
## fp059              -6.721221e-02
## fp060               .
## fp061              -7.760589e-02
## fp062               .
## fp063               5.980146e-02
## fp064               1.178022e-01
## fp065              -6.858867e-02
## fp066               2.027763e-02
## fp067               .
## fp068               9.410473e-05
## fp069               6.202686e-02
## fp070              -4.107573e-02
## fp071               4.439276e-02
## fp072               .
## fp073              -5.244602e-02
## fp074               4.930687e-02
## fp075               8.839814e-02
## fp076               7.493144e-02
## fp077               3.605461e-02
## fp078              -6.189315e-02
## fp079               8.547728e-02
## fp080               .
## fp081              -8.887068e-02
## fp082               6.023673e-02
## fp083              -1.534482e-01
## fp084               1.127010e-01
## fp085              -1.412071e-01
## fp086              -4.139251e-03
## fp087               .
## fp088               4.231616e-02
## fp089               .
## fp090               .
## fp091               2.433162e-04
## fp092               .
## fp093               6.380587e-02
## fp094              -7.055705e-02
## fp095               .
## fp096              -1.801869e-02
## fp097               .
```

```
## fp098              -2.265214e-02
## fp099               7.059742e-02
## fp100               .
## fp101               .
## fp102               4.413964e-05
## fp103              -4.557131e-02
## fp104              -3.418110e-02
## fp105              -2.348530e-02
## fp106               2.708937e-02
## fp107               .
## fp108               .
## fp109               1.213112e-01
## fp110               .
## fp111              -1.391565e-01
## fp112               .
## fp113               4.240678e-02
## fp114               .
## fp115               .
## fp116               1.460550e-02
## fp117               .
## fp118              -3.695554e-02
## fp119               7.398454e-02
## fp120              -2.944697e-03
## fp121               .
## fp122               7.521187e-02
## fp123               .
## fp124               1.127649e-01
## fp125               1.729075e-02
## fp126              -5.432407e-02
## fp127              -1.706544e-01
## fp128              -8.023184e-02
## fp129               .
## fp130              -8.260998e-02
## fp131               6.322328e-02
## fp132              -7.261229e-03
## fp133              -4.992962e-02
## fp134               .
## fp135               6.381773e-02
## fp136               .
## fp137               6.670541e-02
## fp138               7.196695e-02
## fp139               .
## fp140               4.821418e-03
## fp141              -2.868444e-02
## fp142               1.409029e-01
## fp143               8.264788e-02
## fp144               .
## fp145              -2.099669e-02
## fp146               .
## fp147               4.391216e-02
```

```
## fp148              -1.280112e-02
## fp149               .
## fp150               4.742941e-03
## fp151               .
## fp152               .
## fp153               .
## fp154              -9.640976e-02
## fp155               6.192111e-03
## fp156              -5.376009e-02
## fp157              -1.356626e-02
## fp158               .
## fp159               1.432564e-02
## fp160              -1.006388e-02
## fp161              -1.562159e-02
## fp162               .
## fp163               8.844138e-02
## fp164               1.850904e-01
## fp165               .
## fp166               1.015407e-02
## fp167              -4.369410e-02
## fp168               .
## fp169              -5.562562e-02
## fp170               5.799069e-03
## fp171               9.067788e-02
## fp172              -1.928137e-01
## fp173               1.188401e-01
## fp174              -3.622439e-02
## fp175               .
## fp176               1.275764e-01
## fp177               .
## fp178               .
## fp179               .
## fp180              -2.612950e-02
## fp181               5.202021e-02
## fp182              -7.816541e-03
## fp183               .
## fp184               8.485732e-02
## fp185               .
## fp186              -5.209984e-02
## fp187               5.308108e-02
## fp188               5.312134e-02
## fp189               5.624209e-05
## fp190               6.981439e-02
## fp191               1.954827e-02
## fp192               1.600821e-02
## fp193               .
## fp194               .
## fp195               .
## fp196               .
## fp197               .
```

```
## fp198                3.525224e-02
## fp199                .
## fp200                .
## fp201               -6.358390e-02
## fp202                1.809848e-01
## fp203                2.297531e-02
## fp204                .
## fp205                .
## fp206               -1.407785e-02
## fp207                .
## fp208                .
## mol_weight          -6.387481e-01
## num_atoms            .
## num_non_h_atoms      .
## num_bonds            .
## num_non_h_bonds     -8.499806e-01
## num_mult_bonds      -2.419287e-01
## num_rot_bonds       -1.654658e-01
## num_dbl_bonds        .
## num_aromatic_bonds  -9.564355e-02
## num_hydrogen         1.263357e-01
## num_carbon          -6.329914e-01
## num_nitrogen         3.853640e-02
## num_oxygen           3.103962e-01
## num_sulfer          -3.500681e-02
## num_chlorine        -9.542114e-02
## num_halogen          .
## num_rings           -1.869993e-04
## hydrophilic_factor   .
## surface_area1        1.128655e+00
## surface_area2        .
```

```
#test error
pred_lasso <- predict(lasso.fit, test)
mse(test$solubility, pred_lasso)
```

```
## [1] 0.4981467
```

We found the test erorr to be 0.4981467 with a lambda value of 0.004731394.

##Question 1D 1d. Fit a principle component regression model on the training data, with M chosen by cross-validation. Report the test error, along with the value of M selected by cross-validation.

```
ctrl1 <-trainControl(method = "repeatedcv", number = 10, repeats = 5)

set.seed(2)
pcr.fit <-train(x, y,
                method = "pcr",
                tuneGrid  =data.frame(ncomp = 1:228),
                trControl = ctrl1,
```

```
               preProc =c("center", "scale"))
pcr.fit

## Principal Component Analysis
##
## 951 samples
## 228 predictors
##
## Pre-processing: centered (228), scaled (228)
## Resampling: Cross-Validated (10 fold, repeated 5 times)
## Summary of sample sizes: 855, 856, 855, 855, 856, 856, ...
## Resampling results across tuning parameters:
##
##   ncomp  RMSE          Rsquared    MAE
##     1    2.035993e+00  0.01308784  1.575740e+00
##     2    1.971782e+00  0.07486862  1.556857e+00
##     3    1.704827e+00  0.30830677  1.347999e+00
##     4    1.601237e+00  0.38982557  1.244194e+00
##     5    1.572203e+00  0.41146219  1.227212e+00
##     6    1.440871e+00  0.50779073  1.113231e+00
##     7    1.290748e+00  0.60595952  1.004436e+00
##     8    1.288621e+00  0.60729377  1.004671e+00
##     9    1.287914e+00  0.60775694  1.005217e+00
##    10    1.264713e+00  0.62140077  9.790554e-01
##    11    1.241140e+00  0.63557726  9.582758e-01
##    12    1.239767e+00  0.63628107  9.605895e-01
##    13    1.238967e+00  0.63690661  9.588820e-01
##    14    1.190828e+00  0.66436451  9.259003e-01
##    15    1.163083e+00  0.67972533  9.120382e-01
##    16    1.107769e+00  0.70994383  8.693536e-01
##    17    1.050720e+00  0.73841227  8.293929e-01
##    18    1.043546e+00  0.74214773  8.229138e-01
##    19    1.032386e+00  0.74744930  8.094413e-01
##    20    1.011923e+00  0.75759699  7.915013e-01
##    21    1.003308e+00  0.76170176  7.842954e-01
##    22    1.002980e+00  0.76185492  7.840357e-01
##    23    9.780587e-01  0.77357244  7.661997e-01
##    24    9.768092e-01  0.77416789  7.660266e-01
##    25    9.749464e-01  0.77499957  7.649890e-01
##    26    9.689755e-01  0.77765637  7.591298e-01
##    27    9.624417e-01  0.78042435  7.541471e-01
##    28    9.588256e-01  0.78219946  7.508749e-01
##    29    9.573836e-01  0.78273437  7.487453e-01
##    30    9.589455e-01  0.78202684  7.504400e-01
##    31    9.431460e-01  0.78921640  7.348144e-01
##    32    9.259082e-01  0.79693586  7.218993e-01
##    33    9.164310e-01  0.80092851  7.143111e-01
##    34    9.127360e-01  0.80251886  7.116616e-01
##    35    8.989068e-01  0.80821913  7.021523e-01
##    36    8.864354e-01  0.81311625  6.922270e-01
```

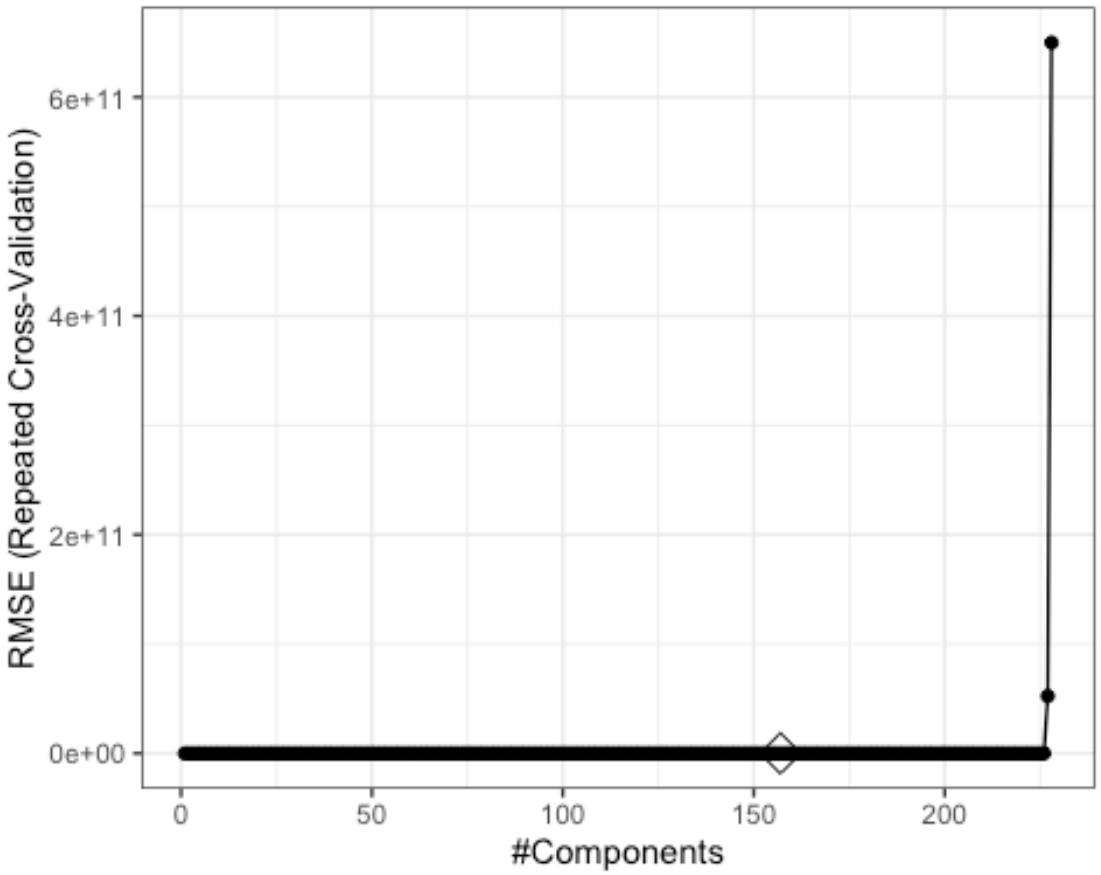
```
##     37    8.826016e-01  0.81495770  6.878036e-01
##     38    8.817588e-01  0.81528240  6.875611e-01
##     39    8.790485e-01  0.81642924  6.853606e-01
##     40    8.741910e-01  0.81832263  6.799648e-01
##     41    8.691057e-01  0.82040135  6.766634e-01
##     42    8.679779e-01  0.82088523  6.753898e-01
##     43    8.637775e-01  0.82270433  6.701415e-01
##     44    8.515971e-01  0.82743660  6.578719e-01
##     45    8.486916e-01  0.82866195  6.527593e-01
##     46    8.476268e-01  0.82899774  6.513548e-01
##     47    8.411880e-01  0.83147359  6.462091e-01
##     48    8.418107e-01  0.83125837  6.463507e-01
##     49    8.417347e-01  0.83126426  6.471674e-01
##     50    8.384376e-01  0.83271407  6.454764e-01
##     51    8.339831e-01  0.83443717  6.427808e-01
##     52    8.303935e-01  0.83577594  6.399474e-01
##     53    8.283483e-01  0.83656039  6.381636e-01
##     54    8.268211e-01  0.83726610  6.363340e-01
##     55    8.265713e-01  0.83740699  6.363466e-01
##     56    8.260894e-01  0.83763934  6.366467e-01
##     57    8.248120e-01  0.83814824  6.352697e-01
##     58    8.219959e-01  0.83927609  6.330002e-01
##     59    8.206073e-01  0.83991497  6.330464e-01
##     60    8.149960e-01  0.84210213  6.277505e-01
##     61    8.066137e-01  0.84536240  6.215989e-01
##     62    8.044109e-01  0.84618527  6.205729e-01
##     63    8.043434e-01  0.84617134  6.207172e-01
##     64    8.039825e-01  0.84636368  6.212087e-01
##     65    8.052967e-01  0.84591795  6.222640e-01
##     66    8.057989e-01  0.84578385  6.220707e-01
##     67    8.061872e-01  0.84562544  6.219988e-01
##     68    8.062301e-01  0.84565739  6.222222e-01
##     69    8.036982e-01  0.84665667  6.205925e-01
##     70    8.033579e-01  0.84686874  6.209219e-01
##     71    8.003580e-01  0.84800040  6.177279e-01
##     72    7.949010e-01  0.85006784  6.134272e-01
##     73    7.932748e-01  0.85067505  6.104269e-01
##     74    7.926590e-01  0.85093637  6.098979e-01
##     75    7.930643e-01  0.85069204  6.097128e-01
##     76    7.930746e-01  0.85077353  6.090697e-01
##     77    7.937645e-01  0.85048004  6.094931e-01
##     78    7.925869e-01  0.85095958  6.083132e-01
##     79    7.913655e-01  0.85140984  6.069831e-01
##     80    7.908993e-01  0.85156969  6.065299e-01
##     81    7.891973e-01  0.85221533  6.053987e-01
##     82    7.892632e-01  0.85219403  6.064309e-01
##     83    7.885800e-01  0.85240428  6.063147e-01
##     84    7.882394e-01  0.85256569  6.053192e-01
##     85    7.864405e-01  0.85319885  6.029727e-01
##     86    7.877500e-01  0.85274755  6.041657e-01
```

```
##    87    7.879005e-01  0.85268084  6.045133e-01
##    88    7.886282e-01  0.85243337  6.050038e-01
##    89    7.877537e-01  0.85282255  6.042942e-01
##    90    7.851212e-01  0.85376162  6.019764e-01
##    91    7.828436e-01  0.85461110  5.999506e-01
##    92    7.805103e-01  0.85552257  5.973381e-01
##    93    7.795797e-01  0.85574917  5.959948e-01
##    94    7.783194e-01  0.85626162  5.948545e-01
##    95    7.788102e-01  0.85605579  5.947215e-01
##    96    7.798598e-01  0.85577729  5.954047e-01
##    97    7.801861e-01  0.85563918  5.951628e-01
##    98    7.817507e-01  0.85510104  5.958956e-01
##    99    7.799884e-01  0.85583851  5.938969e-01
##   100    7.780211e-01  0.85657048  5.920071e-01
##   101    7.764609e-01  0.85703685  5.904360e-01
##   102    7.730304e-01  0.85828007  5.874382e-01
##   103    7.701069e-01  0.85926330  5.848603e-01
##   104    7.667819e-01  0.86045179  5.817471e-01
##   105    7.672152e-01  0.86033690  5.826431e-01
##   106    7.682209e-01  0.86000133  5.837346e-01
##   107    7.666412e-01  0.86059980  5.822010e-01
##   108    7.669629e-01  0.86057536  5.820818e-01
##   109    7.656539e-01  0.86100100  5.807155e-01
##   110    7.640677e-01  0.86153010  5.799071e-01
##   111    7.632600e-01  0.86199722  5.792899e-01
##   112    7.623027e-01  0.86228270  5.785432e-01
##   113    7.632865e-01  0.86197640  5.787320e-01
##   114    7.626153e-01  0.86214359  5.779548e-01
##   115    7.627232e-01  0.86203104  5.779338e-01
##   116    7.618490e-01  0.86233913  5.777116e-01
##   117    7.622110e-01  0.86226450  5.788924e-01
##   118    7.608072e-01  0.86285519  5.773671e-01
##   119    7.593031e-01  0.86333629  5.761589e-01
##   120    7.571041e-01  0.86406945  5.747502e-01
##   121    7.557541e-01  0.86452059  5.747894e-01
##   122    7.540263e-01  0.86522087  5.728905e-01
##   123    7.534109e-01  0.86548143  5.735486e-01
##   124    7.494903e-01  0.86695433  5.704467e-01
##   125    7.461987e-01  0.86815580  5.668017e-01
##   126    7.437004e-01  0.86891547  5.653438e-01
##   127    7.440568e-01  0.86886733  5.659078e-01
##   128    7.445952e-01  0.86856306  5.663020e-01
##   129    7.437654e-01  0.86889142  5.656070e-01
##   130    7.377659e-01  0.87087039  5.605397e-01
##   131    7.351742e-01  0.87176967  5.588278e-01
##   132    7.317804e-01  0.87273750  5.559828e-01
##   133    7.306027e-01  0.87310853  5.547482e-01
##   134    7.312654e-01  0.87301594  5.545017e-01
##   135    7.297403e-01  0.87352208  5.539798e-01
##   136    7.291549e-01  0.87365130  5.532181e-01
```

```
##   137    7.283758e-01  0.87406700  5.528661e-01
##   138    7.283202e-01  0.87411718  5.532537e-01
##   139    7.279363e-01  0.87414839  5.530388e-01
##   140    7.275538e-01  0.87429785  5.527744e-01
##   141    7.255786e-01  0.87501996  5.525626e-01
##   142    7.229023e-01  0.87591896  5.518298e-01
##   143    7.197443e-01  0.87700777  5.508051e-01
##   144    7.160418e-01  0.87825320  5.486352e-01
##   145    7.135979e-01  0.87913905  5.471958e-01
##   146    7.121418e-01  0.87959485  5.460530e-01
##   147    7.117761e-01  0.87974147  5.452900e-01
##   148    7.110782e-01  0.87993125  5.444582e-01
##   149    7.091304e-01  0.88065915  5.426985e-01
##   150    7.088013e-01  0.88074906  5.419025e-01
##   151    7.101359e-01  0.88036167  5.430372e-01
##   152    7.095640e-01  0.88052036  5.425453e-01
##   153    7.105930e-01  0.88015170  5.437342e-01
##   154    7.114918e-01  0.87997196  5.444934e-01
##   155    7.103050e-01  0.88033157  5.441322e-01
##   156    7.089766e-01  0.88079867  5.440627e-01
##   157    7.087392e-01  0.88086786  5.433190e-01
##   158    7.092465e-01  0.88085789  5.439072e-01
##   159    7.102472e-01  0.88056308  5.447390e-01
##   160    7.092047e-01  0.88094446  5.441872e-01
##   161    7.100253e-01  0.88075085  5.446706e-01
##   162    7.113688e-01  0.88040474  5.456977e-01
##   163    7.117708e-01  0.88030142  5.461971e-01
##   164    7.119657e-01  0.88027856  5.461067e-01
##   165    7.115940e-01  0.88029493  5.465168e-01
##   166    7.126067e-01  0.87985263  5.471163e-01
##   167    7.129407e-01  0.87981877  5.471794e-01
##   168    7.138061e-01  0.87956704  5.482239e-01
##   169    7.153799e-01  0.87907663  5.491692e-01
##   170    7.164477e-01  0.87874163  5.506679e-01
##   171    7.184743e-01  0.87806728  5.516881e-01
##   172    7.195169e-01  0.87788637  5.517628e-01
##   173    7.205058e-01  0.87754731  5.517588e-01
##   174    7.190624e-01  0.87803850  5.512242e-01
##   175    7.186607e-01  0.87821582  5.507049e-01
##   176    7.185489e-01  0.87818488  5.511392e-01
##   177    7.184346e-01  0.87824242  5.512443e-01
##   178    7.184625e-01  0.87817525  5.516731e-01
##   179    7.182806e-01  0.87822659  5.521802e-01
##   180    7.185207e-01  0.87814736  5.519802e-01
##   181    7.195515e-01  0.87780159  5.524797e-01
##   182    7.194989e-01  0.87783551  5.522199e-01
##   183    7.201029e-01  0.87768023  5.517631e-01
##   184    7.197335e-01  0.87780226  5.504070e-01
##   185    7.188020e-01  0.87801762  5.499548e-01
##   186    7.176109e-01  0.87843934  5.493310e-01
```

```
##   187    7.185002e-01  0.87811760  5.493281e-01
##   188    7.188429e-01  0.87797000  5.491299e-01
##   189    7.206429e-01  0.87740149  5.494793e-01
##   190    7.205444e-01  0.87738460  5.491915e-01
##   191    7.212349e-01  0.87718819  5.492680e-01
##   192    7.242844e-01  0.87632045  5.516116e-01
##   193    7.243560e-01  0.87625382  5.514795e-01
##   194    7.272494e-01  0.87531918  5.523067e-01
##   195    7.223728e-01  0.87680688  5.495499e-01
##   196    7.210395e-01  0.87716400  5.490867e-01
##   197    7.201758e-01  0.87743675  5.487954e-01
##   198    7.204637e-01  0.87735439  5.480159e-01
##   199    7.206957e-01  0.87715193  5.472094e-01
##   200    7.241377e-01  0.87601270  5.488776e-01
##   201    7.277955e-01  0.87486072  5.509193e-01
##   202    7.249506e-01  0.87586808  5.487549e-01
##   203    7.270697e-01  0.87531058  5.498610e-01
##   204    7.293697e-01  0.87466577  5.513687e-01
##   205    7.292654e-01  0.87470389  5.515927e-01
##   206    7.245903e-01  0.87633032  5.488279e-01
##   207    7.238424e-01  0.87654516  5.484378e-01
##   208    7.212574e-01  0.87731429  5.483324e-01
##   209    7.216098e-01  0.87701782  5.478025e-01
##   210    7.240234e-01  0.87624589  5.488911e-01
##   211    7.270770e-01  0.87529032  5.503707e-01
##   212    7.271297e-01  0.87517307  5.500668e-01
##   213    7.298136e-01  0.87422537  5.515348e-01
##   214    7.290412e-01  0.87453241  5.514688e-01
##   215    7.295591e-01  0.87432256  5.515691e-01
##   216    7.277437e-01  0.87501823  5.497012e-01
##   217    7.277570e-01  0.87514073  5.497463e-01
##   218    7.282231e-01  0.87500236  5.485786e-01
##   219    7.280177e-01  0.87513362  5.485776e-01
##   220    7.287780e-01  0.87496287  5.493692e-01
##   221    7.312232e-01  0.87437447  5.504454e-01
##   222    7.311069e-01  0.87447764  5.513130e-01
##   223    7.291215e-01  0.87529785  5.504792e-01
##   224    7.284889e-01  0.87545198  5.503401e-01
##   225    7.194392e-01  0.87846073  5.459468e-01
##   226    7.204280e-01  0.87820488  5.462547e-01
##   227    5.239989e+10  0.86145965  7.602975e+09
##   228    6.499149e+11  0.72658669  6.707725e+10
##
## RMSE was used to select the optimal model using the smallest value.
## The final value used for the model was ncomp = 157.

#M value
pcr.fit$bestTune
```

```
##     ncomp
## 157   157

#test error
pred_pcr <- predict(pcr.fit, test)
mse(test$solubility, pred_pcr)

## [1] 0.549917

ggplot(pcr.fit, highlight = TRUE) + theme_bw()
```



We found the test erorr to be 0.549917 with an M value of 157 selected by cross-validation.

##Question 1E 1e. Briefly discuss the results obtained in (a)~(d).

In Questions A - D, we fit four models using the training data and calculated the mean square error using the test data. Our data set includes 228 predictor variables. Our outcome is solubility (the solubility of a compound). We fit four different models (linear regression, ridge regression, lasso, and PCR) using repeated cross validation to determine which model would fit the data best. We measured the mean squared error (MSE) to quantify the extent to which the predicted reponse value for a given observation is close to the true response value for that observation. In general, the smaller the MSE is, the closer the predicted responses are to the true responses.

First, we fit a linear model using least squares on all the predictors in the training data. We found that the MSE calculated on the test data was 0.5558898.

Next, we fit two models on all the predictor variables using two different techniques that "shrink" the coefficient estimates towards zero, which reduces variance. We fit a ridge regression model on the training data. Alpha was held at a value of 0 and our final lambda value chosen by cross-validation was 0.0800381. Using our test data, we found that test error was 0.5134603. Next, we fit a lasso model on the training data. Alpha was held at a value of 1 and our final lambda value chosen by cross-validation was 0.004731394. Our test error was 0.4981467.

Lastly, we fit a principle component regression (PCR) model on the training data with M chosen by cross validation. The PCR method constructs the first M principal components and then uses the componenets as the predictors in a linear regression model. Our M-value was 157 and our test error was 0.549917.

##Question 1F 1f. Which model will you choose for predicting solubility?

```
resamp <- resamples(list(lm = lmfit,
                         ridge = ridge.fit,
                         lasso = lasso.fit,
                         pcr = pcr.fit))

summary(resamp)

##
## Call:
## summary.resamples(object = resamp)
##
## Models: lm, ridge, lasso, pcr
## Number of resamples: 50
##
## MAE
##            Min.   1st Qu.    Median      Mean   3rd Qu.      Max. NA's
## lm    0.4151018 0.5042577 0.5288620 0.5304540 0.5607729 0.6838928    0
## ridge 0.4236514 0.4901614 0.5218728 0.5213242 0.5473545 0.6319687    0
## lasso 0.4241879 0.4841574 0.5182456 0.5172918 0.5485782 0.6376831    0
## pcr   0.4213811 0.5066631 0.5508642 0.5433190 0.5726849 0.6668606    0
##
## RMSE
##            Min.   1st Qu.    Median      Mean   3rd Qu.      Max. NA's
## lm    0.5763256 0.6702141 0.7006166 0.7093576 0.7470323 0.9234601    0
## ridge 0.5547163 0.6488957 0.6786781 0.6856755 0.7234002 0.8333606    0
## lasso 0.5573310 0.6414382 0.6693349 0.6774968 0.7143975 0.8315571    0
## pcr   0.5763791 0.6725372 0.7162175 0.7087392 0.7458388 0.8730497    0
##
## Rsquared
##            Min.   1st Qu.    Median      Mean   3rd Qu.      Max. NA's
## lm    0.8105782 0.8668429 0.8844072 0.8814378 0.8966132 0.9313096    0
## ridge 0.8326964 0.8764872 0.8908486 0.8880644 0.9001630 0.9264558    0
```

```
## lasso 0.8375038 0.8776310 0.8945815 0.8904842 0.9045490 0.9282464    0
## pcr   0.8131551 0.8632151 0.8848280 0.8808679 0.8975444 0.9275388    0
```

The model I would choose for predicting solubility is lasso because it has the smallest mean RMSE value of 0.6774968. Next, I would choose ridge with an RSME value of 0.6856755, then I would choose PCR with an RMSE value of 0.7087392, and last I would choose the linear model with an RSME value of 0.7093576.