Is624 Assignment3

James Quacinella

06/22/2015

Contents

Question 7.2

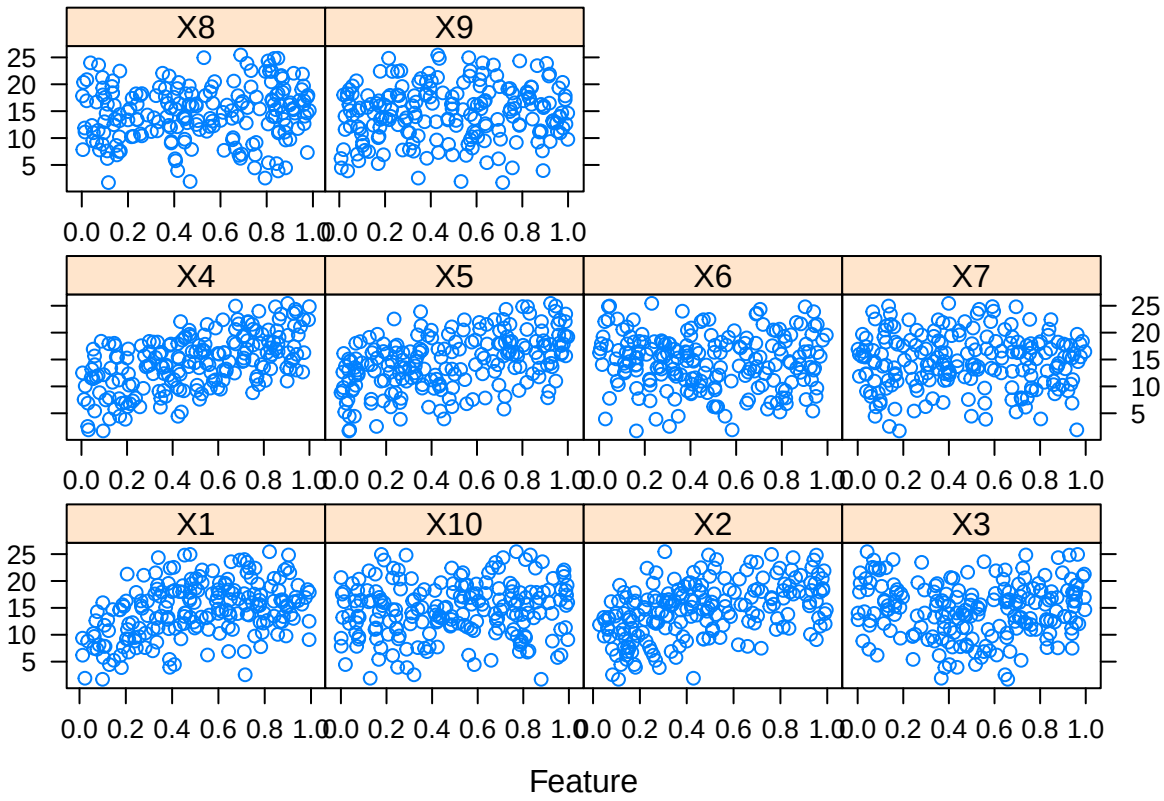
Question

Friedman (1991) introduced several benchmark data sets create by simulation. One of these simulations used the following nonlinear equation to create data:

$$y = 10sin(nx_1x_2) + 20(x_3 - 0.5)2 + 10x_4 + 5x_5 + N(0, alpha^2)$$

where the x values are random variables uniformly distributed between [0, 1] (there are also 5 other non-informative variables also created in the simula- tion). The package mlbench contains a function called mlbench.friedman1 that simulates these data:

```
trainingData <- mlbench.friedman1(200, sd = 1)
## We convert the ' x ' data from a matrix to a data
## frame One reason is that this will give the
## columns names.
trainingData$x <- data.frame(trainingData$x)
## Look at the data using
featurePlot(trainingData$x, trainingData$y)
```




```
## or other methods.  This creates a list with a
## vector ' y ' and a matrix of predictors ' x ' .
## Also simulate a large test set to estimate the
## true error rate with good precision:
testData <- mlbench.friedman1(5000, sd = 1)
testData$x <- data.frame(testData$x)
```

Tune several models on these data. For example:

```
knnModel <- train(x = trainingData$x, y = trainingData$y,
    method = "knn", preProc = c("center", "scale"),
    tuneLength = 10)

knnModel
```

```
## k-Nearest Neighbors
##
## 200 samples
##  10 predictors
##
## Pre-processing: centered, scaled
## Resampling: Bootstrapped (25 reps)
##
## Summary of sample sizes: 200, 200, 200, 200, 200, 200, ...
##
## Resampling results across tuning parameters:
##
##   k   RMSE      Rsquared   RMSE SD    Rsquared SD
```

```
##    5  3.492611  0.5514878  0.1890666  0.05654488
##    7  3.367499  0.5919769  0.1854113  0.05577892
##    9  3.309609  0.6182302  0.2022743  0.05494413
##   11  3.289578  0.6367648  0.2270144  0.04463323
##   13  3.281397  0.6559647  0.2285598  0.04659637
##   15  3.287363  0.6661678  0.2404266  0.04807383
##   17  3.334535  0.6641188  0.2509047  0.05246330
##   19  3.370126  0.6642655  0.2527395  0.05007358
##   21  3.403700  0.6655961  0.2556742  0.05093303
##   23  3.435312  0.6667585  0.2498849  0.05135945
##
## RMSE was used to select the optimal model using  the smallest value.
## The final value used for the model was k = 13.
```

```
knnPred <- predict(knnModel, newdata = testData$x)
## The function ' postResample ' can be used to get
## the test set perforamnce values
postResample(pred = knnPred, obs = testData$y)
```

```
##      RMSE  Rsquared
## 3.1454316 0.6661258
```

Answer

A kNN model was generated for us above. We have three options: train a neural network, a MARS model, and a SVM model (with potentially different kernels). Lets first look at a neural net model.

Neural Net Model

```
# Create a Neural Net model
nnetFit <- nnet(trainingData$x, trainingData$y, size = 5,
    decay = 0.01, linout = TRUE, trace = FALSE, maxit = 500,
    MaxNWts = 5 * (ncol(trainingData$x) + 1) + 5 +
        1)

# Show the fit
nnetFit
```

```
## a 10-5-1 network with 61 weights
## options were - linear output units  decay=0.01
```

```
summary(nnetFit)
```

```
## a 10-5-1 network with 61 weights
## options were - linear output units  decay=0.01
##   b->h1  i1->h1  i2->h1  i3->h1  i4->h1  i5->h1  i6->h1  i7->h1  i8->h1
##   -0.87   -0.97   -0.67    1.04    1.20    0.50    0.00    0.12    0.03
##  i9->h1 i10->h1
##    0.08   -0.15
##   b->h2  i1->h2  i2->h2  i3->h2  i4->h2  i5->h2  i6->h2  i7->h2  i8->h2
##    0.70   -0.94   -0.56   -6.14    0.33    0.16    0.41    0.16    0.42
##  i9->h2 i10->h2
##   -0.13    0.09
##   b->h3  i1->h3  i2->h3  i3->h3  i4->h3  i5->h3  i6->h3  i7->h3  i8->h3
##   -4.22    3.73   12.72    1.24   -0.99   -0.10    0.07   -0.69   -0.51
##  i9->h3 i10->h3
##   -0.19    0.40
##   b->h4  i1->h4  i2->h4  i3->h4  i4->h4  i5->h4  i6->h4  i7->h4  i8->h4
##    0.02   -7.52    0.88   -1.00    0.16   -0.03    0.23    0.26    0.74
##  i9->h4 i10->h4
##    0.42   -0.08
##   b->h5  i1->h5  i2->h5  i3->h5  i4->h5  i5->h5  i6->h5  i7->h5  i8->h5
##   -8.50    7.68    7.82   -0.09   -0.40    0.44   -0.31    0.12   -0.11
##  i9->h5 i10->h5
##    0.13    0.70
##   b->o  h1->o  h2->o  h3->o  h4->o  h5->o
## -12.65  37.16  18.40   6.72 -11.46   9.49
```

```
# Predict on the test set and see how well it did
nnetPredict <- predict(nnetFit, testData$x)
postResample(pred = nnetPredict, obs = testData$y)
```

```
##      RMSE  Rsquared
## 1.4423934 0.9167035
```

Discussion: This basic model ends up with a high R^2 with no tuning of any model parameters. Lets see what we can do with better tuning.

TODO: Grid search params using train like example in book

MARS Model

```
# Create a MARS model
marsFit <- earth(trainingData$x, trainingData$y)

# Show the fit
marsFit
```

```
## Selected 12 of 20 terms, and 7 of 10 predictors
## Termination condition: Reached nk 21
## Importance: X4, X2, X1, X5, X3, X10, X8, X6-unused, X7-unused, ...
## Number of terms at each degree of interaction: 1 11 (additive model)
## GCV 2.802104    RSS 438.9356    GRSq 0.8927606    RSq 0.9151612
```

```
summary(marsFit)
```

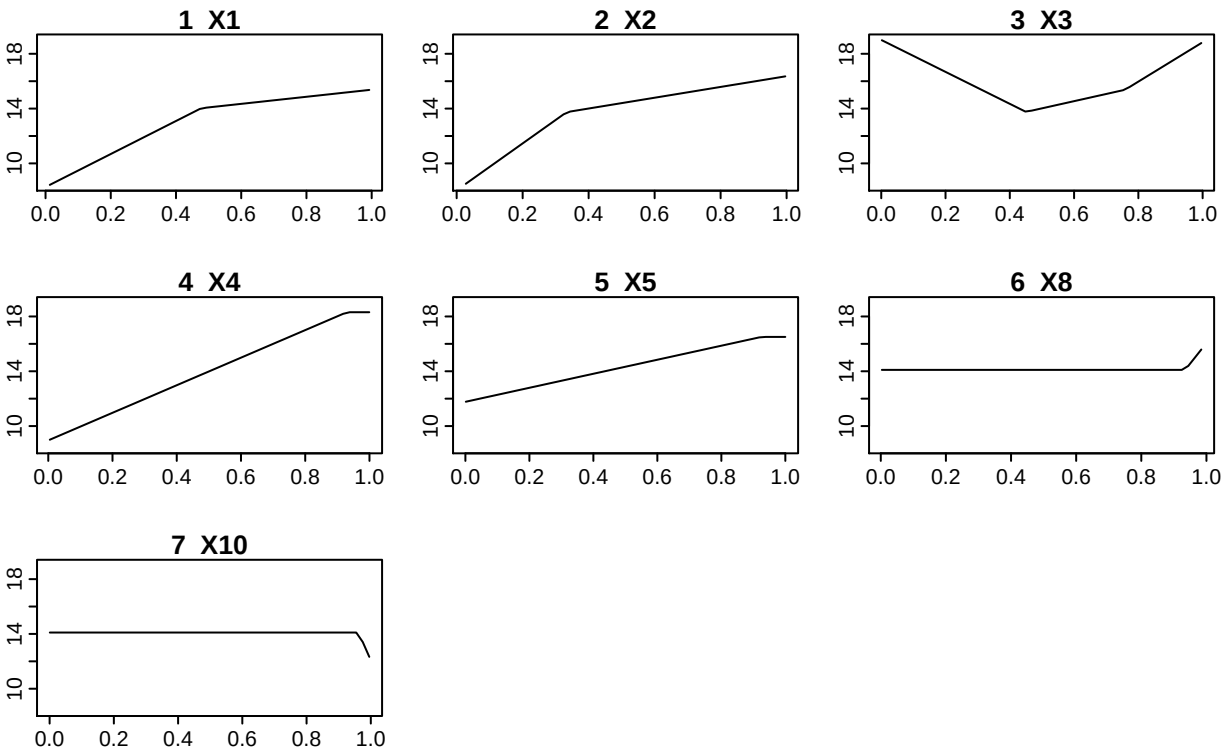
```
## Call: earth(x=trainingData$x, y=trainingData$y)
##
##                 coefficients
## (Intercept)        19.942097
## h(0.47762-X1)     -12.049125
## h(X1-0.47762)       2.584661
## h(0.333521-X2)    -17.108629
## h(X2-0.333521)      3.938851
## h(0.450485-X3)     11.661496
## h(X3-0.450485)      5.244239
## h(X3-0.758697)      9.131904
## h(0.929743-X4)    -10.060992
## h(0.925323-X5)     -5.119769
## h(X8-0.934716)     29.960793
## h(X10-0.961949)   -53.215152
##
## Selected 12 of 20 terms, and 7 of 10 predictors
## Termination condition: Reached nk 21
## Importance: X4, X2, X1, X5, X3, X10, X8, X6-unused, X7-unused, ...
## Number of terms at each degree of interaction: 1 11 (additive model)
## GCV 2.802104    RSS 438.9356    GRSq 0.8927606    RSq 0.9151612
```

```
# Predict on the test set and see how well it did
marsPredict <- predict(marsFit, testData$x)
postResample(pred = marsPredict, obs = testData$y)
```

```
##      RMSE  Rsquared
## 1.8871553 0.8598973
```

```
# Inspect the model's use of predictors
plotmo(marsFit, caption = "Predictors vs Observed in Additive MARS Model")
```

```
##  grid:    X1        X2        X3        X4      X5        X6        X7
##    0.5054907 0.4244831 0.5158712 0.5114502 0.45547 0.4441768 0.4696555
##         X8        X9       X10
##  0.4949538 0.4917092 0.5036008
```



Discussion: It does seem like the first 5 predictors are noted as important, and left the other predictors out. Pretty impressive there, MARS.

Support Vector Machine Model

```
svmFit <- ksvm(y ~ ., data = as.data.frame(trainingData),
    kernel = "rbfdot", kpar = "automatic", C = 1, epsilon = 0.1)
```

```
## Using automatic sigma estimation (sigest) for RBF or laplace kernel
```

```
svmFit
```

```
## Support Vector Machine object of class "ksvm"
##
## SV type: eps-svr  (regression)
##  parameter : epsilon = 0.1  cost C = 1
##
## Gaussian Radial Basis kernel function.
##  Hyperparameter : sigma =  0.0617877542444368
##
## Number of Support Vectors : 170
##
## Objective Function Value : -42.2388
## Training error : 0.083273
```

```
summary(svmFit)
```

```
## Length  Class   Mode
##      1   ksvm     S4
```

```
svmPredict <- predict(svmFit, as.data.frame(testData))
postResample(pred = svmPredict, obs = testData$y)
```

```
##      RMSE  Rsquared
## 2.4201570 0.7770665
```

Lets tune a hopefuly better SVM model:

```
svmRTuned <- train(trainingData$x, trainingData$y,
    method = "svmRadial", preProc = c("center", "scale"),
    tuneLength = 14, trControl = trainControl(method = "cv"))

svmRTuned
```

```
## Support Vector Machines with Radial Basis Function Kernel
##
## 200 samples
##  10 predictors
##
## Pre-processing: centered, scaled
## Resampling: Cross-Validated (10 fold)
##
```

```
## Summary of sample sizes: 180, 180, 180, 180, 180, 180, ...
##
## Resampling results across tuning parameters:
##
##   C        RMSE      Rsquared   RMSE SD    Rsquared SD
##      0.25  2.595022  0.7929425  0.3592234  0.07841429
##      0.50  2.359706  0.8062457  0.3528681  0.08495943
##      1.00  2.249207  0.8183889  0.3751124  0.07884515
##      2.00  2.177391  0.8291890  0.3386386  0.06241932
##      4.00  2.133625  0.8354071  0.3228081  0.05105185
##      8.00  2.105195  0.8388782  0.2884709  0.04641157
##     16.00  2.104702  0.8389732  0.2840710  0.04436779
##     32.00  2.104702  0.8389732  0.2840710  0.04436779
##     64.00  2.104702  0.8389732  0.2840710  0.04436779
##    128.00  2.104702  0.8389732  0.2840710  0.04436779
##    256.00  2.104702  0.8389732  0.2840710  0.04436779
##    512.00  2.104702  0.8389732  0.2840710  0.04436779
##   1024.00  2.104702  0.8389732  0.2840710  0.04436779
##   2048.00  2.104702  0.8389732  0.2840710  0.04436779
##
## Tuning parameter 'sigma' was held constant at a value of 0.0664479
## RMSE was used to select the optimal model using  the smallest value.
## The final values used for the model were sigma = 0.0664479 and C = 16.
```

```
summary(svmRTuned)
```

```
## Length  Class   Mode
##      1   ksvm     S4
```

```
svmTunedPredict <- predict(svmRTuned, testData$x)
postResample(pred = svmTunedPredict, obs = testData$y)
```

```
##      RMSE  Rsquared
## 1.9861165 0.8427245
```

TODO: Try other kernels

Question 7.4

Question

Return to the permeability problem outlined in Exercise 6.2. Train several nonlinear regression models and evaluate the resampling and test set performance.

(a) Which nonlinear regression model gives the optimal resampling and test set performance?

(b) Do any of the nonlinear models outperform the optimal linear model you previously developed in Exercise 6.2? If so, what might this tell you about the underlying relationship between the predictors and the response?

(c) Would you recommend any of the models you have developed to replace the permeability laboratory experiment?

Answer

Like above, we will try out three kinds of models, and we'll see which one does the best and if these models are better than the linear models created in the previous assignment.

Data Pre-Processing

Before creating models, we should make sure the data is properly pre-processed. I will follow the same exact moethodology as per the previous assignment:

```
# Get rid of any predictors that are nero-zero
# variance
nearZero <- nearZeroVar(fingerprints)
fingerprints.filtered <- fingerprints[, -nearZero]

# Filter out highly correlated predictors
correlations <- cor(fingerprints.filtered)
highCorr <- findCorrelation(correlations, cutoff = 0.9)
fingerprints.filtered <- fingerprints.filtered[, -highCorr]

# Split the data into a training and test set
fingerprints.train <- fingerprints.filtered[1:124,
    ]
fingerprints.test <- fingerprints.filtered[1:124, ]
permeability.train <- permeability[1:124, ]
permeability.test <- permeability[1:124, ]
```

The three models will be presented next, with results to follow.

Neural Net Model

```
# Create a Neural Net model
nnetFit <- nnet(fingerprints.train, permeability.train,
    size = 5, decay = 0.01, linout = TRUE, trace = FALSE,
    maxit = 500, MaxNWts = 5 * (ncol(fingerprints.test) +
        1) + 5 + 1)

# Show the fit
nnetFit
```

```
## a 110-5-1 network with 561 weights
## options were - linear output units  decay=0.01
```

```
# summary(nnetFit)

# Predict on the test set and see how well it did
nnetPredict <- predict(nnetFit, fingerprints.test)
postResample(pred = nnetPredict, obs = permeability.test)
```

```
##      RMSE  Rsquared
## 2.2985847 0.9779633
```

Discussion:

MARS Model

```
# Create a MARS model
marsFit <- earth(fingerprints.train, permeability.train)

# Show the fit
marsFit
```

```
## Selected 16 of 55 terms, and 15 of 110 predictors
## Termination condition: GRSq -10 at 55 terms
## Importance: X157, X503, X698-unused, X6-unused, X141-unused, X371, ...
## Number of terms at each degree of interaction: 1 15 (additive model)
## GCV 113.6165    RSS 7924.748    GRSq 0.5333509    RSq 0.7332244
```

```
summary(marsFit)
```

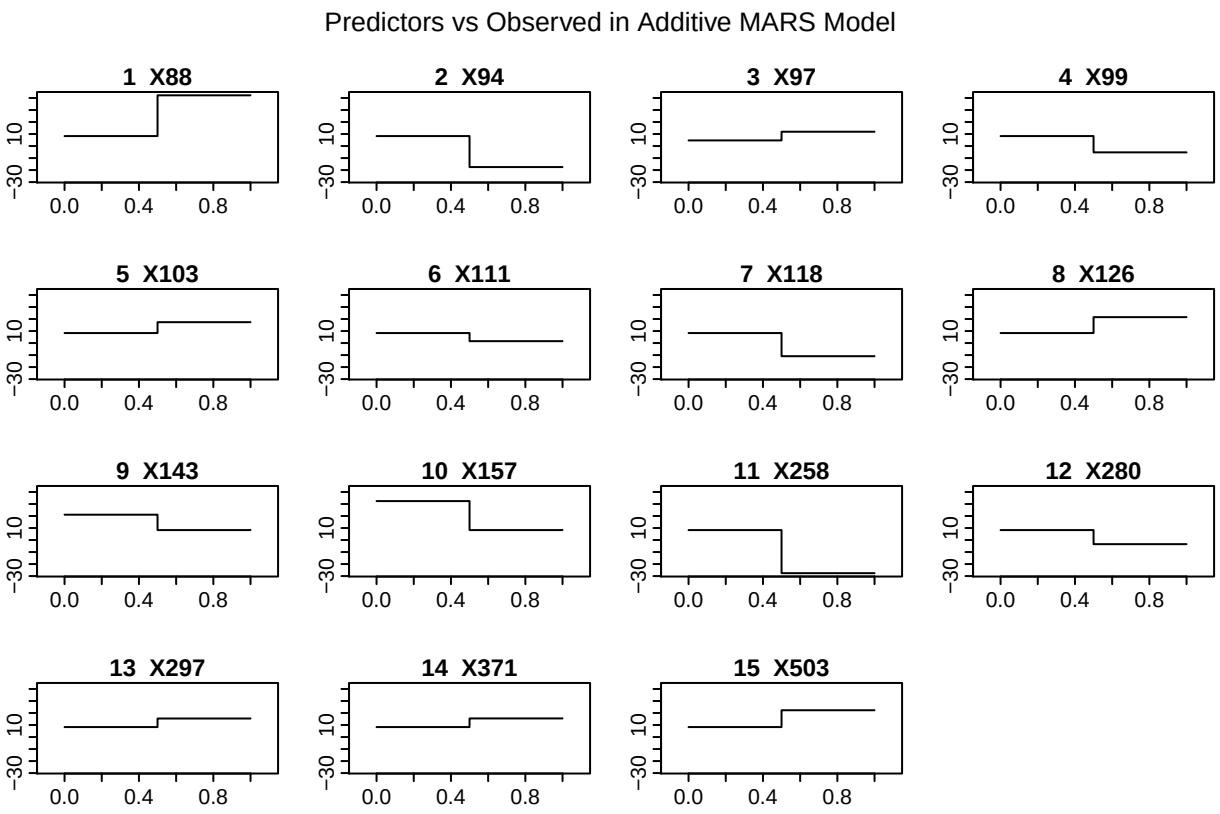
```
## Call: earth(x=fingerprints.train, y=permeability.train)
##
##             coefficients
## (Intercept)    41.602269
## X88            33.911235
## X94           -25.793638
## X97             7.231244
## X99           -13.560189
## X103            9.060141
## X111           -6.731119
## X118          -19.261123
## X126           13.264522
## X143          -12.805263
## X157          -24.116205
## X258          -35.873602
## X280          -11.735595
## X297            7.127412
## X371            7.197599
## X503           13.991710
##
## Selected 16 of 55 terms, and 15 of 110 predictors
## Termination condition: GRSq -10 at 55 terms
## Importance: X157, X503, X698-unused, X6-unused, X141-unused, X371, ...
## Number of terms at each degree of interaction: 1 15 (additive model)
## GCV 113.6165    RSS 7924.748    GRSq 0.5333509    RSq 0.7332244
```

```
# Predict on the test set and see how well it did
marsPredict <- predict(marsFit, fingerprints.test)
postResample(pred = marsPredict, obs = permeability.test)
```

```
##      RMSE  Rsquared
## 7.9943264 0.7332244
```

```r
# Inspect the model's use of predictors
plotmo(marsFit, caption = "Predictors vs Observed in Additive MARS Model")
```

```
##  grid:    X1  X6 X11 X12 X15 X16 X25 X35 X36 X41 X48 X86 X88 X93 X94 X96
##            0 0.5   0   1   0   0   0   0   0   0   0   1   0   0   0   1
##  X97 X98 X99 X103 X111 X118 X121 X125 X126 X141 X143 X146 X157 X158 X159
##  0.5   0   0    0    0    0    0    0    0    0    1    0    1    1    1
##  X182 X221 X225 X226 X229 X230 X231 X235 X237 X238 X242 X251 X257 X258
##     1    1    0    1    1    0    1    1    1    0    0    0    1    0
##  X260 X262 X264 X267 X272 X276 X278 X280 X293 X295 X296 X297 X306 X310
##     1    1    0    1    0    0    0    0    0    0    0    0    0    0
##  X314 X315 X316 X319 X329 X334 X337 X338 X340 X345 X356 X357 X358 X361
##     0    1    0    0    0    0    0    0    0    0    0    0    0    0
##  X362 X366 X371 X372 X374 X377 X499 X503 X507 X508 X509 X510 X511 X512
##     0    0    0    0    0    0    0    0    0    0    0    0    0    0
##  X514 X519 X520 X549 X551 X561 X568 X571 X573 X577 X590 X598 X602 X613
##     0    0    0    0    0    0    0    0    0    0    0    0    0    0
##  X679 X698 X699 X700 X732 X750 X780 X795 X812
##     0    0    0    0    0    0    0    0    0
```



Discussion:

Support Vector Machine Model

```
df.fingerprints.train <- as.data.frame(fingerprints.train)
df.fingerprints.train$y <- permeability.train
df.fingerprints.test <- as.data.frame(fingerprints.test)

svmFit <- ksvm(y ~ ., data = df.fingerprints.train,
    kernel = "rbfdot", kpar = "automatic", C = 1, epsilon = 0.1)
```

```
## Using automatic sigma estimation (sigest) for RBF or laplace kernel
```

```
svmFit
```

```
## Support Vector Machine object of class "ksvm"
##
## SV type: eps-svr  (regression)
##  parameter : epsilon = 0.1  cost C = 1
##
## Gaussian Radial Basis kernel function.
##  Hyperparameter : sigma =  0.00538822397317305
##
## Number of Support Vectors : 98
##
## Objective Function Value : -37.555
## Training error : 0.277195
```

```
summary(svmFit)
```

```
## Length  Class   Mode
##      1   ksvm     S4
```

```
svmPredict <- predict(svmFit, df.fingerprints.test)
postResample(pred = svmPredict, obs = permeability.test)
```

```
##      RMSE  Rsquared
## 8.1820126 0.7476109
```

Lets tune a hopefuly better SVM model:

```
svmRTuned <- train(fingerprints.train, permeability.train,
    method = "svmRadial", preProc = c("center", "scale"),
    tuneLength = 14, trControl = trainControl(method = "cv"))

svmRTuned
```

```
## Support Vector Machines with Radial Basis Function Kernel
##
## 124 samples
## 110 predictors
```

```
##
## Pre-processing: centered, scaled
## Resampling: Cross-Validated (10 fold)
##
## Summary of sample sizes: 112, 110, 112, 112, 112, 112, ...
##
## Resampling results across tuning parameters:
##
##   C        RMSE       Rsquared   RMSE SD   Rsquared SD
##      0.25  12.027841  0.6067874  3.592147  0.2240992
##      0.50  10.930337  0.6194370  3.741335  0.2158559
##      1.00  10.268580  0.6253968  3.722541  0.1988653
##      2.00  10.204321  0.6290063  3.322751  0.1991679
##      4.00   9.924063  0.6314036  2.895832  0.1719067
##      8.00  10.025663  0.6132986  2.505033  0.1429011
##     16.00  10.414601  0.5824339  2.194583  0.1364803
##     32.00  10.560171  0.5735247  2.091467  0.1389528
##     64.00  10.678761  0.5620420  2.085183  0.1505723
##    128.00  10.678881  0.5620398  2.085497  0.1505819
##    256.00  10.678850  0.5620488  2.085614  0.1506134
##    512.00  10.678888  0.5620333  2.085875  0.1506271
##   1024.00  10.678715  0.5620446  2.085703  0.1506189
##   2048.00  10.679173  0.5620031  2.086108  0.1506230
##
## Tuning parameter 'sigma' was held constant at a value of 0.00510706
## RMSE was used to select the optimal model using  the smallest value.
## The final values used for the model were sigma = 0.00510706 and C = 4.
```

```
summary(svmRTuned)
```

```
## Length  Class   Mode
##      1   ksvm     S4
```

```
svmTunedPredict <- predict(svmRTuned, fingerprints.test)
postResample(pred = svmTunedPredict, obs = permeability.test)
```

```
##      RMSE  Rsquared
## 4.5165979 0.9189726
```

Discussion:

Results and Discussions