Chapter 8,9

Chapter 8. Regression Trees and Rule-Based Models

Advantages of tree-based & rule-based models: - Effectively handle many types of predictors (sparse, skewed, continuous, categorical, etc) without the need to pre-process them - Effectively handle missing data and implicitly conduct feature selection

Disadvantages: - Model instability - Less-than-optimal predictive performance

8.1 Basic regression trees

Regression trees determine: The predictor to split on and value of the split The depth or complexity of the tree The prediction equation in the terminal nodes

Classification and regression tree (CART)

Recursive partitioning

Prune cost-complexity tuning

Conditional inference regression trees

8.2 Regression model trees

One limitation of simple regression trees is that each terminal node uses the average of the training set outcomes in that node for prediction. As a consequence, these models may not do a good job predicting samples whose true outcomes are extremely high or low.

Model tree: - The splitting criterion is different - The terminal nodes predict the outcome using a linear model (as opposed to the simple average) - When a sample is predicted, it is often a combination of the predictions from different models along the same path through the tree

Once the tree is fully grown, there is a linear model for every node in the tree

Smoothing and pruning

8.3 Rule-Based models

8.4 Bagged trees

Bagging is a general approach that uses bootstrapping in conjunction with any regression or classification model to construct an ensemble. Each model in the ensemble is then used to generate a prediction for a new sample and these m predictions are averaged to give the bagged model's prediction

Bagging effectively reduced the variance of a prediction through its aggregation process

Bagging can provide their own internal estimate of predictive performance that correlates well with either cross-validation estimates or test set estimates

Parameter: the number of bootstrap samples to aggregate, m

The most imporvement in prediction performance is obtained with a small number of trees ($m < 10$)

Disadvantages: computational costs and memory requirements increase as the number of bootstrap samples increases (need parallel computing); A bagged model is much less interpretable than a model that is not bagged

8.5 Random forests

Tree correlation problem

Parameter: the number of randomly selected predictors, k, to choose from at each split; the number of trees for the forest

Quantify the impact of predictors in the ensemble: variable importance scores

8.6 Boosting

Gradient boosting machines: given a loss function (e.g. squared error for regression) and a weak learner (e.g. regression trees), the algorithm seeks to find an addictive model that minimizes the loss function. The algorithm is typically initialized with the best guess of the response (e.g. the mean of the response in regression). The gradient (e.g residual) is calculated, and a model is then fit to the residuals to minimize the loss function. The current model is added to the previous model, and the procedure continues for a user-specified number of iterations.

Two parameters: tree depth (interaction depth) and number of iterations

In random forests, all trees are created independently, each tree is created to have maximum depth, and each tree contributes equally to the final model. The trees in boosting, however, are dependent on past trees, have minimum depth, and contribute unequally to the final model

Parameter: learning rate lambda

Stochastic gradient boosting: parameter: the fraction of training data used

8.7 Cubist

8.8 Computing

```r
library(AppliedPredictiveModeling)
data(solubility)
```

Single trees

```r
library(rpart)
library(party)
```

```
## Loading required package: grid

## Loading required package: mvtnorm

## Loading required package: modeltools

## Loading required package: stats4

## Loading required package: strucchange

## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

## Loading required package: sandwich
```

```
library(caret)
```

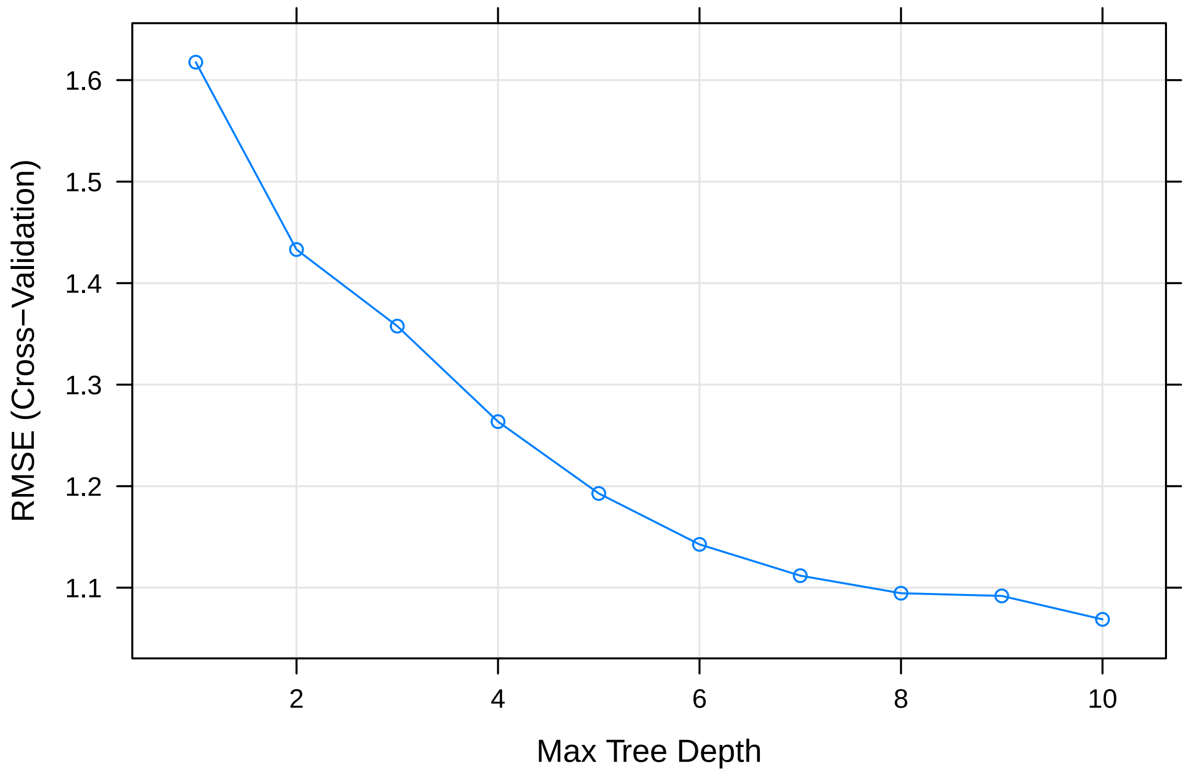
```
## Loading required package: lattice

## Loading required package: ggplot2
```

```
# rpartTree <- rpart(y ~. , data = trainData)
# ctreeTree <- ctree(y ~., data = trainData)
```

```
set.seed(100)
rpartTune <- train(solTrainXtrans, solTrainY,
                   method = "rpart2", tuneLength = 10, trControl = trainControl(method = "cv"))
rpartTune
```

```
## CART
##
## 951 samples
## 228 predictors
##
## No pre-processing
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 856, 856, 855, 855, 857, 856, ...
## Resampling results across tuning parameters:
##
##   maxdepth  RMSE      Rsquared   MAE
##    1        1.617667  0.3745252  1.2657915
##    2        1.433114  0.5067404  1.1326186
##    3        1.357672  0.5568291  1.0657348
##    4        1.263596  0.6166997  0.9974476
##    5        1.192831  0.6581800  0.9429124
##    6        1.142654  0.6853056  0.9009065
##    7        1.111858  0.7020728  0.8707216
##    8        1.094535  0.7110088  0.8545809
##    9        1.091880  0.7116190  0.8465921
##   10        1.068799  0.7236716  0.8232469
##
## RMSE was used to select the optimal model using the smallest value.
## The final value used for the model was maxdepth = 10.
```

```
plot(rpartTune)
```



```r
library(partykit)
```

```
## Loading required package: libcoin

##
## Attaching package: 'partykit'

## The following objects are masked from 'package:party':
##
##     cforest, ctree, ctree_control, edge_simple, mob, mob_control,
##     node_barplot, node_bivplot, node_boxplot, node_inner,
##     node_surv, node_terminal, varimp
```

```r
# rpartTree2 <- as.party(rpartTree)
# plot(rpartTree2)
```

Model trees

```r
# library(rJava)
# library(RWeka)
# m5tree <- M5P(y ~., data = trainData)
# m5rules <- M5Rules(y ~., data = trainData)

# m5tree <- M5P(y~., data = trainData, control = Weka_control(M = 10))
```

```r
set.seed(100)
# m5Tune <- train(solTrainXtrans, solTrainY,
#                  method = "M5", trControl = trainControl(method = "cv"),
                # use an option for M5() to specify the minimum number of samples needed to further spl
                # to be 10
#                  control = Weka_control(M = 10))
# m5Tune
```

Bagged trees

```r
library(ipred)
# baggedTree <- ipredbagg(solTrainY, soltrainXtrans)
# baggedTree <- bagging(y ~., data = trainData)
```

```r
library(party)
# The mtry parameter should be the number of predictors (the number of columns minus 1 for the outcome)
# bagCtrl <- cforest_control(mtry = ncol(trainData) - 1)
# baggedTree <- cforest(y ~., data = trainData, controls = bagCtrl)
```

Random forest

```r
library(randomForest)
```

```
## randomForest 4.6-14
```

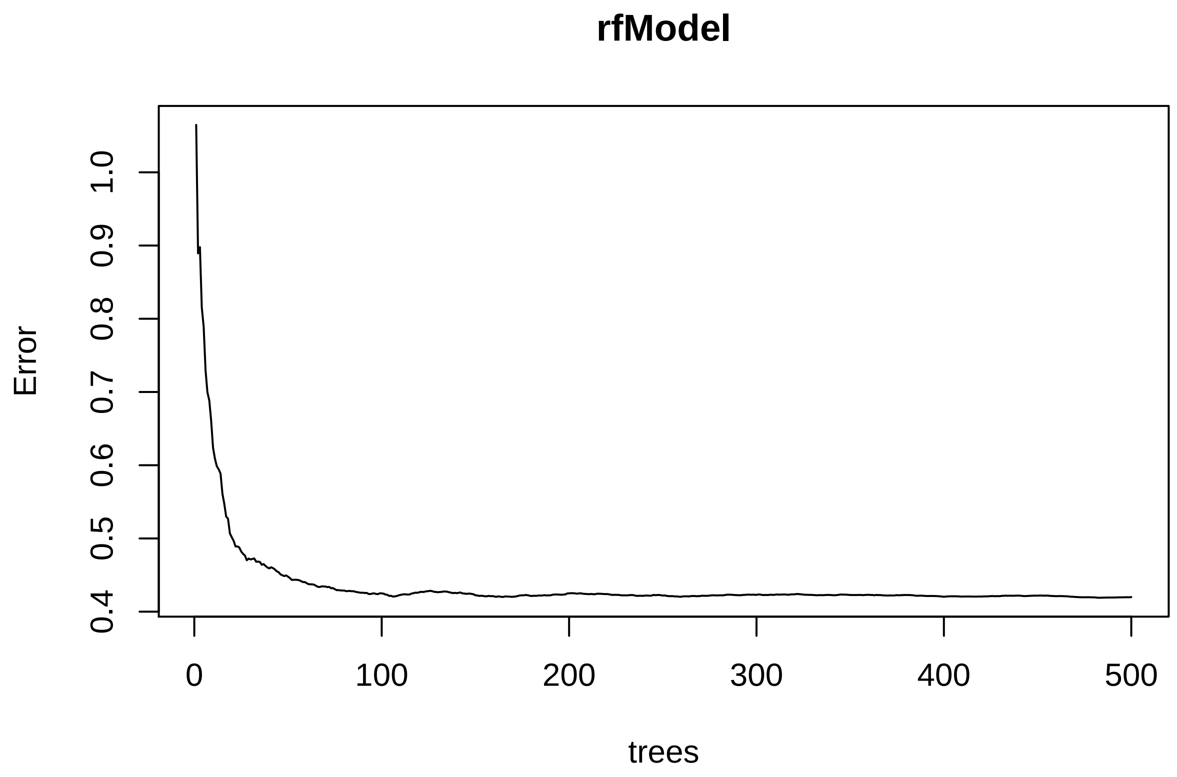
```
## Type rfNews() to see new features/changes/bug fixes.
```

```
##
## Attaching package: 'randomForest'
```

```
## The following object is masked from 'package:ggplot2':
##
##     margin
```

```r
# rfModel <- randomForest(solTrainXtrans, solTrainY)
# plot(rfModel)
# rfModel <- randomForest(y~., data = trainData)
```

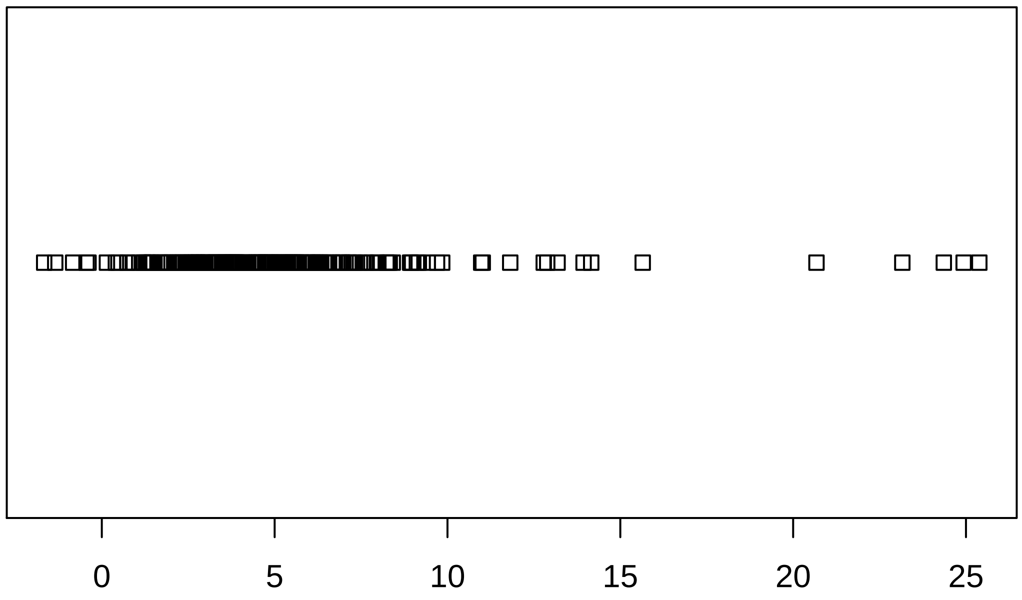
```r
rfModel <- randomForest(solTrainXtrans, solTrainY, importance = TRUE, ntrees = 1000)
plot(rfModel)
```



```r
varImp(rfModel)
```

```
##                         Overall
## FP001               4.9035397
## FP002               4.5173336
## FP003               5.2651449
## FP004               5.3534272
## FP005               4.0012304
## FP006               4.5216878
## FP007               5.1426926
## FP008               3.7664504
## FP009               8.9419800
## FP010               4.0913812
## FP011               7.1621673
## FP012               2.5938663
## FP013               5.9640441
## FP014               5.5458559
## FP015               5.6063018
## FP016               1.3534664
## FP017               2.3103872
## FP018               4.8878891
## FP019               4.8081074
## FP020               1.0819293
## FP021               4.9369350
## FP022               0.4811916
## FP023               3.1273097
## FP024               4.7322551
## FP025               2.9871009
## FP026               4.4669480
## FP027              1.8015534
## FP028              6.2350386
## FP029              4.5928359
## FP030              1.9402305
## FP031             -0.4428272
## FP032              1.8647345
## FP033              2.3096049
## FP034              2.1833253
## FP035              2.7599828
## FP036              1.6779065
## FP037              2.6913617
## FP038              2.1412629
## FP039              4.2427324
## FP040              6.2455952
## FP041              2.7764855
## FP042              3.0578549
## FP043              6.6773979
## FP044              3.8763379
## FP045             -1.3474740
## FP046              5.6391848
## FP047              4.1450616
## FP048              5.2275021
## FP049              5.0494072
## FP050              5.1152601
## FP051              4.7850243
## FP052              2.4203972
## FP053              3.2778305
## FP054              3.3908081
## FP055              4.3767248
## FP056              3.2960228
## FP057              3.7357665
## FP058              4.4015344
## FP059              8.2398287
## FP060              3.6518753
## FP061              3.9035034
## FP062              5.1725711
## FP063              8.4166648
## FP064              6.0390521
## FP065              9.4279484
## FP066              4.8519458
## FP067              3.9540626
## FP068              5.7774273
## FP069              3.7847642
## FP070              7.9861729
## FP071              4.7435926
## FP072              8.9174516
## FP073              3.9104699
## FP074              5.5100927
## FP075              9.1138152
## FP076              4.7442442
## FP077              4.5246009
## FP078              4.7270107
## FP079              3.9169103
## FP080              4.7781063
## FP081                 5.9099012
## FP082                 6.8824461
## FP083                 6.1204186
## FP084                 7.2772927
## FP085                 5.7758982
## FP086                 2.9210661
## FP087                 3.4824032
## FP088                 6.0892663
## FP089                 4.1402188
## FP090                 3.6968301
## FP091                 3.8642820
## FP092                 8.3293599
## FP093                 3.5172756
## FP094                 6.2820149
## FP095                 6.9447058
## FP096                 5.3819609
## FP097                 5.2009166
## FP098                 3.5706170
## FP099                 4.5917802
## FP100                 4.0714470
## FP101                 7.9568773
## FP102                 5.4663399
## FP103                 5.6309280
## FP104                 8.2349019
## FP105                 6.0990519
## FP106                 5.2413898
## FP107                 6.5747172
## FP108                 3.6798614
## FP109                 3.2232942
## FP110                 4.3204147
## FP111                 4.9214950
## FP112                 7.4685953
## FP113                 3.8298512
## FP114                 2.5548354
## FP115                 3.5425867
## FP116                 9.8471206
## FP117                 2.6831705
## FP118                 3.0693098
## FP119                 3.3274056
## FP120                 2.7751091
## FP121                 3.7185092
## FP122                 4.2386276
## FP123                 3.2021216
## FP124                 5.5815977
## FP125                 5.7707275
## FP126                 3.8904296
## FP127                 3.0748570
## FP128                 1.1928388
## FP129                 1.4335981
## FP130                 1.3247069
## FP131                 7.3237994
## FP132                 3.5978795
## FP133                 0.7408498
## FP134                 6.3207675
## FP135             6.4879687
## FP136             5.3161813
## FP137             6.9852551
## FP138             5.1828949
## FP139             2.2393345
## FP140             0.4068018
## FP141             6.6738775
## FP142             6.8688696
## FP143             2.5891762
## FP144             0.9273491
## FP145             3.4015914
## FP146             2.9252752
## FP147             6.9608496
## FP148             7.7979962
## FP149             3.4881913
## FP150             0.7405528
## FP151             1.8918487
## FP152             3.7777972
## FP153             3.8376854
## FP154             3.4866155
## FP155             3.6500487
## FP156             1.2210742
## FP157            -0.8290406
## FP158             1.4338128
## FP159             3.3689630
## FP160             5.9592501
## FP161             4.5947157
## FP162             6.3467641
## FP163             6.4296641
## FP164             5.5248303
## FP165             4.6477029
## FP166             3.4063255
## FP167             5.5163883
## FP168             6.5195618
## FP169             6.1898410
## FP170             4.1633035
## FP171             4.4543863
## FP172             4.8018051
## FP173             4.4365747
## FP174             2.2447256
## FP175             5.4433280
## FP176             5.6883289
## FP177             2.3303875
## FP178             5.6668162
## FP179             5.5353319
## FP180             2.7041966
## FP181             2.5405726
## FP182            -1.6671453
## FP183             1.3958681
## FP184             3.7390860
## FP185             2.4724426
## FP186             1.8220840
## FP187             3.2581056
## FP188             3.8026551
## FP189              3.4974055
## FP190              2.7880770
## FP191              2.0180086
## FP192              2.7841783
## FP193              2.1054090
## FP194              2.9734030
## FP195              3.0631626
## FP196              0.1472486
## FP197              1.4641090
## FP198              5.1092625
## FP199              1.2818414
## FP200              1.6902729
## FP201              2.9502826
## FP202              8.2859369
## FP203              2.6447558
## FP204              7.1324342
## FP205             -0.3849690
## FP206              7.2251395
## FP207              5.0061713
## FP208              1.7056865
## MolWeight         24.9329131
## NumAtoms          12.8850551
## NumNonHAtoms      14.1581379
## NumBonds          11.8154320
## NumNonHBonds      13.1844644
## NumMultBonds      11.0202211
## NumRotBonds       15.6480178
## NumDblBonds        9.6986176
## NumAromaticBonds   7.6628249
## NumHydrogen       13.9375314
## NumCarbon         20.6748643
## NumNitrogen        9.1703789
## NumOxygen         10.9748840
## NumSulfer          7.6285450
## NumChlorine        8.9766358
## NumHalogen        12.7848440
## NumRings           5.1188278
## HydrophilicFactor 24.3589002
## SurfaceArea1      25.3857678
## SurfaceArea2      23.1582602
```

```
plot(varImp(rfModel))
```



Boosted trees

```
library(gbm)
```

```
## Loaded gbm 2.1.5
```

```
# gbmModel <- gbm.fit(solTrainXtrans, solTrainY, distribution = "gaussian")
# gbmModel <- gbm(y~., data = trainData, distribution = "gaussian")
```

```
gbmGrid <- expand.grid(.interaction.depth=seq(1,7,by=2), .n.trees = seq(100, 1000, by = 50),
                       .shrinkage = c(0.01, 0.1), .n.minobsinnode = c(1,2))
set.seed(100)
gbmTune <- train(solTrainXtrans, solTrainY, method = "gbm", tuneGrid = gbmGrid,
                 # The gbm() function produces copious amounts of output, so pass in the verbose option
                 # printing a lot to the screen
                 verbose = FALSE)
```

```
gbmTune
```

```
## Stochastic Gradient Boosting
##
## 951 samples
## 228 predictors
##
## No pre-processing
## Resampling: Bootstrapped (25 reps)
## Summary of sample sizes: 951, 951, 951, 951, 951, 951, ...
## Resampling results across tuning parameters:
##
##   shrinkage  interaction.depth  n.minobsinnode  n.trees  RMSE
##   0.01       1                  1                100     1.5861368
##    0.01        1                  1                 150      1.4478299
##    0.01        1                  1                 200      1.3420154
##    0.01        1                  1                 250      1.2606514
##    0.01        1                  1                 300      1.1976185
##    0.01        1                  1                 350      1.1471647
##    0.01        1                  1                 400      1.1069772
##    0.01        1                  1                 450      1.0736272
##    0.01        1                  1                 500      1.0454466
##    0.01        1                  1                 550      1.0217346
##    0.01        1                  1                 600      1.0004965
##    0.01        1                  1                 650      0.9813886
##    0.01        1                  1                 700      0.9642774
##    0.01        1                  1                 750      0.9487750
##    0.01        1                  1                 800      0.9344848
##    0.01        1                  1                 850      0.9216602
##    0.01        1                  1                 900      0.9100551
##    0.01        1                  1                 950      0.8992110
##    0.01        1                  1                1000      0.8892561
##    0.01        1                  2                 100      1.5861091
##    0.01        1                  2                 150      1.4482286
##    0.01        1                  2                 200      1.3425701
##    0.01        1                  2                 250      1.2615685
##    0.01        1                  2                 300      1.1984495
##    0.01        1                  2                 350      1.1477524
##    0.01        1                  2                 400      1.1066458
##    0.01        1                  2                 450      1.0734638
##    0.01        1                  2                 500      1.0452368
##    0.01        1                  2                 550      1.0212606
##    0.01        1                  2                 600      0.9999803
##    0.01        1                  2                 650      0.9813084
##    0.01        1                  2                 700      0.9637068
##    0.01        1                  2                 750      0.9481956
##    0.01        1                  2                 800      0.9342978
##    0.01        1                  2                 850      0.9219089
##    0.01        1                  2                 900      0.9098243
##    0.01        1                  2                 950      0.8992693
##    0.01        1                  2                1000      0.8894152
##    0.01        3                  1                 100      1.3306383
##    0.01        3                  1                 150      1.1671771
##    0.01        3                  1                 200      1.0592781
##    0.01        3                  1                 250      0.9843150
##    0.01        3                  1                 300      0.9297331
##    0.01        3                  1                 350      0.8895160
##    0.01        3                  1                 400      0.8577882
##    0.01        3                  1                 450      0.8325357
##    0.01        3                  1                 500      0.8120476
##    0.01        3                  1                 550      0.7949200
##    0.01        3                  1                 600      0.7806953
##    0.01        3                  1                 650      0.7687084
##    0.01        3                  1                 700      0.7583977
##    0.01        3                  1                 750      0.7490910
##    0.01        3                  1                 800      0.7412874
##    0.01        3                  1                 850      0.7342660
##    0.01        3                  1                 900      0.7280522
##    0.01         3                   1                950      0.7226449
##    0.01         3                   1               1000      0.7177714
##    0.01         3                   2                100      1.3297369
##    0.01         3                   2                150      1.1668258
##    0.01         3                   2                200      1.0588547
##    0.01         3                   2                250      0.9833292
##    0.01         3                   2                300      0.9290514
##    0.01         3                   2                350      0.8883771
##    0.01         3                   2                400      0.8572820
##    0.01         3                   2                450      0.8324423
##    0.01         3                   2                500      0.8121587
##    0.01         3                   2                550      0.7953541
##    0.01         3                   2                600      0.7810967
##    0.01         3                   2                650      0.7691135
##    0.01         3                   2                700      0.7587014
##    0.01         3                   2                750      0.7500616
##    0.01         3                   2                800      0.7421810
##    0.01         3                   2                850      0.7353346
##    0.01         3                   2                900      0.7297065
##    0.01         3                   2                950      0.7241828
##    0.01         3                   2               1000      0.7192391
##    0.01         5                   1                100      1.2416682
##    0.01         5                   1                150      1.0717773
##    0.01         5                   1                200      0.9642799
##    0.01         5                   1                250      0.8913925
##    0.01         5                   1                300      0.8407019
##    0.01         5                   1                350      0.8038202
##    0.01         5                   1                400      0.7771028
##    0.01         5                   1                450      0.7566434
##    0.01         5                   1                500      0.7402000
##    0.01         5                   1                550      0.7267807
##    0.01         5                   1                600      0.7165220
##    0.01         5                   1                650      0.7076350
##    0.01         5                   1                700      0.7003065
##    0.01         5                   1                750      0.6938448
##    0.01         5                   1                800      0.6879761
##    0.01         5                   1                850      0.6831212
##    0.01         5                   1                900      0.6788508
##    0.01         5                   1                950      0.6751197
##    0.01         5                   1               1000      0.6716461
##    0.01         5                   2                100      1.2426352
##    0.01         5                   2                150      1.0720154
##    0.01         5                   2                200      0.9640632
##    0.01         5                   2                250      0.8908860
##    0.01         5                   2                300      0.8397543
##    0.01         5                   2                350      0.8032035
##    0.01         5                   2                400      0.7763674
##    0.01         5                   2                450      0.7552256
##    0.01         5                   2                500      0.7389990
##    0.01         5                   2                550      0.7257635
##    0.01         5                   2                600      0.7149401
##    0.01         5                   2                650      0.7062292
##    0.01         5                   2                700      0.6989968
##    0.01         5                   2                750      0.6926783
##    0.01         5                  2                800      0.6873110
##    0.01         5                  2                850      0.6825241
##    0.01         5                  2                900      0.6781763
##    0.01         5                  2                950      0.6741406
##    0.01         5                  2               1000      0.6706988
##    0.01         7                  1                100      1.1978624
##    0.01         7                  1                150      1.0216964
##    0.01         7                  1                200      0.9131839
##    0.01         7                  1                250      0.8427074
##    0.01         7                  1                300      0.7945718
##    0.01         7                  1                350      0.7608025
##    0.01         7                  1                400      0.7363311
##    0.01         7                  1                450      0.7183637
##    0.01         7                  1                500      0.7049150
##    0.01         7                  1                550      0.6938266
##    0.01         7                  1                600      0.6853488
##    0.01         7                  1                650      0.6780762
##    0.01         7                  1                700      0.6719439
##    0.01         7                  1                750      0.6671771
##    0.01         7                  1                800      0.6630064
##    0.01         7                  1                850      0.6591254
##    0.01         7                  1                900      0.6558965
##    0.01         7                  1                950      0.6527606
##    0.01         7                  1               1000      0.6500513
##    0.01         7                  2                100      1.1983876
##    0.01         7                  2                150      1.0214227
##    0.01         7                  2                200      0.9130297
##    0.01         7                  2                250      0.8422676
##    0.01         7                  2                300      0.7940181
##    0.01         7                  2                350      0.7607676
##    0.01         7                  2                400      0.7368132
##    0.01         7                  2                450      0.7192380
##    0.01         7                  2                500      0.7056869
##    0.01         7                  2                550      0.6946030
##    0.01         7                  2                600      0.6859324
##    0.01         7                  2                650      0.6790544
##    0.01         7                  2                700      0.6729634
##    0.01         7                  2                750      0.6680199
##    0.01         7                  2                800      0.6636946
##    0.01         7                  2                850      0.6597577
##    0.01         7                  2                900      0.6565093
##    0.01         7                  2                950      0.6536717
##    0.01         7                  2               1000      0.6511138
##    0.10         1                  1                100      0.8911968
##    0.10         1                  1                150      0.8246583
##    0.10         1                  1                200      0.7914973
##    0.10         1                  1                250      0.7705461
##    0.10         1                  1                300      0.7561170
##    0.10         1                  1                350      0.7452299
##    0.10         1                  1                400      0.7366983
##    0.10         1                  1                450      0.7305196
##    0.10         1                  1                500      0.7247545
##    0.10         1                  1                550      0.7212568
##    0.10         1                  1                600      0.7174213
##    0.10       1                   1             650      0.7140293
##    0.10       1                   1             700      0.7105844
##    0.10       1                   1             750      0.7090533
##    0.10       1                   1             800      0.7075326
##    0.10       1                   1             850      0.7057958
##    0.10       1                   1             900      0.7045380
##    0.10       1                   1             950      0.7042939
##    0.10       1                   1            1000      0.7027125
##    0.10       1                   2             100      0.8894093
##    0.10       1                   2             150      0.8253343
##    0.10       1                   2             200      0.7908548
##    0.10       1                   2             250      0.7718040
##    0.10       1                   2             300      0.7572600
##    0.10       1                   2             350      0.7459302
##    0.10       1                   2             400      0.7379919
##    0.10       1                   2             450      0.7308675
##    0.10       1                   2             500      0.7254325
##    0.10       1                   2             550      0.7223353
##    0.10       1                   2             600      0.7184450
##    0.10       1                   2             650      0.7145869
##    0.10       1                   2             700      0.7128162
##    0.10       1                   2             750      0.7124597
##    0.10       1                   2             800      0.7103985
##    0.10       1                   2             850      0.7083554
##    0.10       1                   2             900      0.7065558
##    0.10       1                   2             950      0.7055321
##    0.10       1                   2            1000      0.7049691
##    0.10       3                   1             100      0.7340647
##    0.10       3                   1             150      0.6999913
##    0.10       3                   1             200      0.6809643
##    0.10       3                   1             250      0.6714038
##    0.10       3                   1             300      0.6643582
##    0.10       3                   1             350      0.6606567
##    0.10       3                   1             400      0.6581209
##    0.10       3                   1             450      0.6559008
##    0.10       3                   1             500      0.6538558
##    0.10       3                   1             550      0.6517715
##    0.10       3                   1             600      0.6515053
##    0.10       3                   1             650      0.6508634
##    0.10       3                   1             700      0.6507329
##    0.10       3                   1             750      0.6506738
##    0.10       3                   1             800      0.6512349
##    0.10       3                   1             850      0.6511228
##    0.10       3                   1             900      0.6512849
##    0.10       3                   1             950      0.6511935
##    0.10       3                   1            1000      0.6510366
##    0.10       3                   2             100      0.7240478
##    0.10       3                   2             150      0.6911787
##    0.10       3                   2             200      0.6743931
##    0.10       3                   2             250      0.6648988
##    0.10       3                   2             300      0.6594942
##    0.10       3                   2             350      0.6552437
##    0.10       3                   2             400      0.6527701
##    0.10       3                   2             450      0.6507268
##    0.10       3                  2               500    0.6495829
##    0.10       3                  2               550    0.6486697
##    0.10       3                  2               600    0.6486379
##    0.10       3                  2               650    0.6484445
##    0.10       3                  2               700    0.6479483
##    0.10       3                  2               750    0.6481852
##    0.10       3                  2               800    0.6484753
##    0.10       3                  2               850    0.6485826
##    0.10       3                  2               900    0.6479823
##    0.10       3                  2               950    0.6482080
##    0.10       3                  2              1000    0.6486943
##    0.10       5                  1               100    0.6846223
##    0.10       5                  1               150    0.6645685
##    0.10       5                  1               200    0.6559562
##    0.10       5                  1               250    0.6511906
##    0.10       5                  1               300    0.6488119
##    0.10       5                  1               350    0.6480968
##    0.10       5                  1               400    0.6465298
##    0.10       5                  1               450    0.6464022
##    0.10       5                  1               500    0.6463781
##    0.10       5                  1               550    0.6458390
##    0.10       5                  1               600    0.6458243
##    0.10       5                  1               650    0.6458617
##    0.10       5                  1               700    0.6461748
##    0.10       5                  1               750    0.6459561
##    0.10       5                  1               800    0.6457024
##    0.10       5                  1               850    0.6459412
##    0.10       5                  1               900    0.6457360
##    0.10       5                  1               950    0.6457174
##    0.10       5                  1              1000    0.6458996
##    0.10       5                  2               100    0.6864237
##    0.10       5                  2               150    0.6681927
##    0.10       5                  2               200    0.6590493
##    0.10       5                  2               250    0.6549373
##    0.10       5                  2               300    0.6517903
##    0.10       5                  2               350    0.6514359
##    0.10       5                  2               400    0.6502983
##    0.10       5                  2               450    0.6493628
##    0.10       5                  2               500    0.6487524
##    0.10       5                  2               550    0.6486131
##    0.10       5                  2               600    0.6487598
##    0.10       5                  2               650    0.6487491
##    0.10       5                  2               700    0.6489547
##    0.10       5                  2               750    0.6489449
##    0.10       5                  2               800    0.6495176
##    0.10       5                  2               850    0.6498933
##    0.10       5                  2               900    0.6499328
##    0.10       5                  2               950    0.6499233
##    0.10       5                  2              1000    0.6499125
##    0.10       7                  1               100    0.6673344
##    0.10       7                  1               150    0.6545505
##    0.10       7                  1               200    0.6487791
##    0.10       7                  1               250    0.6460946
##    0.10       7                  1               300    0.6443002
##   0.10       7                  1                  350      0.6439449
##   0.10       7                  1                  400      0.6436185
##   0.10       7                  1                  450      0.6435328
##   0.10       7                  1                  500      0.6433097
##   0.10       7                  1                  550      0.6436604
##   0.10       7                  1                  600      0.6437565
##   0.10       7                  1                  650      0.6446443
##   0.10       7                  1                  700      0.6448751
##   0.10       7                  1                  750      0.6450187
##   0.10       7                  1                  800      0.6454780
##   0.10       7                  1                  850      0.6455668
##   0.10       7                  1                  900      0.6461791
##   0.10       7                  1                  950      0.6465378
##   0.10       7                  1                 1000      0.6462621
##   0.10       7                  2                  100      0.6692586
##   0.10       7                  2                  150      0.6570274
##   0.10       7                  2                  200      0.6512989
##   0.10       7                  2                  250      0.6493463
##   0.10       7                  2                  300      0.6485139
##   0.10       7                  2                  350      0.6483918
##   0.10       7                  2                  400      0.6485717
##   0.10       7                  2                  450      0.6483915
##   0.10       7                  2                  500      0.6483280
##   0.10       7                  2                  550      0.6489457
##   0.10       7                  2                  600      0.6485279
##   0.10       7                  2                  650      0.6484867
##   0.10       7                  2                  700      0.6487945
##   0.10       7                  2                  750      0.6492002
##   0.10       7                  2                  800      0.6496395
##   0.10       7                  2                  850      0.6496315
##   0.10       7                  2                  900      0.6497175
##   0.10       7                  2                  950      0.6494501
##   0.10       7                  2                 1000      0.6497737
##   Rsquared   MAE
##   0.6162707  1.2174887
##   0.6679547  1.1134875
##   0.6985288  1.0327435
##   0.7185520  0.9686027
##   0.7333768  0.9178733
##   0.7451073  0.8777845
##   0.7549246  0.8456924
##   0.7640243  0.8191087
##   0.7721047  0.7967238
##   0.7793779  0.7779925
##   0.7859725  0.7613929
##   0.7919610  0.7465565
##   0.7974477  0.7332452
##   0.8024076  0.7211780
##   0.8067878  0.7100755
##   0.8107027  0.7005150
##   0.8142934  0.6917056
##   0.8177298  0.6836398
##   0.8208502  0.6761528
##   0.6156196  1.2176729
##   0.6686211  1.1143686
##   0.6995087  1.0331801
##   0.7183536  0.9697236
##   0.7334873  0.9190409
##   0.7455043  0.8785850
##   0.7559294  0.8459972
##   0.7646767  0.8197443
##   0.7728366  0.7971850
##   0.7799201  0.7779921
##   0.7864806  0.7612745
##   0.7922027  0.7464527
##   0.7977911  0.7327645
##   0.8026101  0.7207793
##   0.8069334  0.7100263
##   0.8106955  0.7003293
##   0.8144515  0.6912165
##   0.8177460  0.6831574
##   0.8207215  0.6757396
##   0.7375136  1.0277437
##   0.7637376  0.8949234
##   0.7855345  0.8059256
##   0.8025482  0.7454475
##   0.8161746  0.7035104
##   0.8268676  0.6735613
##   0.8355802  0.6503800
##   0.8427877  0.6320948
##   0.8486050  0.6172970
##   0.8535447  0.6049547
##   0.8577443  0.5944600
##   0.8612677  0.5855895
##   0.8643672  0.5780582
##   0.8672419  0.5712342
##   0.8696596  0.5653740
##   0.8718378  0.5599850
##   0.8737992  0.5551303
##   0.8755010  0.5507195
##   0.8770839  0.5468335
##   0.7385378  1.0277181
##   0.7638367  0.8949414
##   0.7858645  0.8056298
##   0.8031114  0.7451290
##   0.8165082  0.7030939
##   0.8271829  0.6731040
##   0.8356936  0.6504763
##   0.8426861  0.6326183
##   0.8485246  0.6179796
##   0.8534797  0.6058404
##   0.8576259  0.5953427
##   0.8612219  0.5864582
##   0.8643467  0.5787511
##   0.8669798  0.5722111
##   0.8694058  0.5661747
##   0.8715273  0.5607307
##   0.8732589  0.5562275
##   0.8750187  0.5518240
##   0.8766034  0.5478019
##   0.7826113  0.9610152
##   0.8031241  0.8209707
##   0.8208214  0.7324585
##   0.8353893  0.6751716
##   0.8466765  0.6369148
##   0.8555126  0.6099010
##   0.8621748  0.5905558
##   0.8675889  0.5752284
##   0.8720839  0.5630787
##   0.8758631  0.5531684
##   0.8787582  0.5450965
##   0.8813514  0.5380755
##   0.8835177  0.5321418
##   0.8854572  0.5268535
##   0.8872189  0.5219850
##   0.8886657  0.5178690
##   0.8899728  0.5142846
##   0.8911267  0.5111495
##   0.8921736  0.5080831
##   0.7824839  0.9628391
##   0.8033197  0.8216187
##   0.8212603  0.7326125
##   0.8357289  0.6748952
##   0.8469577  0.6361943
##   0.8556832  0.6095248
##   0.8624027  0.5901942
##   0.8680202  0.5744177
##   0.8724671  0.5621541
##   0.8762015  0.5521850
##   0.8793139  0.5437064
##   0.8818354  0.5367360
##   0.8839426  0.5309188
##   0.8857973  0.5257122
##   0.8874222  0.5213738
##   0.8888700  0.5174184
##   0.8901913  0.5137307
##   0.8914114  0.5103104
##   0.8924577  0.5074389
##   0.8038910  0.9275847
##   0.8229155  0.7821659
##   0.8390877  0.6934351
##   0.8517106  0.6381406
##   0.8614700  0.6023817
##   0.8691412  0.5772358
##   0.8751095  0.5589149
##   0.8796817  0.5454474
##   0.8832218  0.5351226
##   0.8862597  0.5264514
##   0.8886164  0.5195024
##   0.8907073  0.5134702
##   0.8924803  0.5083141
##   0.8938663  0.5042205
##   0.8950790  0.5005142
##   0.8962306  0.4971301
##   0.8971783  0.4942090
##   0.8981069  0.4912972
##   0.8989077  0.4888387
##   0.8040898  0.9286265
##   0.8233046  0.7817991
##   0.8391974  0.6931980
##   0.8518580  0.6385052
##   0.8616660  0.6019299
##   0.8691348  0.5769831
##   0.8748943  0.5591674
##   0.8793312  0.5458336
##   0.8829422  0.5354422
##   0.8859974  0.5267366
##   0.8884228  0.5198475
##   0.8903650  0.5142235
##   0.8921327  0.5092230
##   0.8935795  0.5051457
##   0.8948558  0.5015429
##   0.8960206  0.4983480
##   0.8969997  0.4955007
##   0.8978452  0.4929719
##   0.8985881  0.4906105
##   0.8187058  0.6778022
##   0.8394731  0.6281717
##   0.8505075  0.6029349
##   0.8577467  0.5866520
##   0.8628586  0.5755670
##   0.8667058  0.5667074
##   0.8698620  0.5594688
##   0.8719283  0.5541545
##   0.8739694  0.5494351
##   0.8752218  0.5463242
##   0.8765606  0.5428113
##   0.8776468  0.5397177
##   0.8788412  0.5368751
##   0.8793905  0.5352507
##   0.8800126  0.5335589
##   0.8805459  0.5322347
##   0.8809458  0.5305326
##   0.8810659  0.5299225
##   0.8816031  0.5281723
##   0.8194538  0.6774188
##   0.8393034  0.6285151
##   0.8507518  0.6021582
##   0.8573277  0.5868563
##   0.8624969  0.5754776
##   0.8666019  0.5670778
##   0.8693150  0.5602613
##   0.8717783  0.5540356
##   0.8736662  0.5496609
##   0.8747267  0.5465991
##   0.8761778  0.5433480
##   0.8774424  0.5397885
##   0.8780771  0.5381771
##   0.8782147  0.5369879
##   0.8789762  0.5349343
##   0.8795939  0.5333000
##   0.8802573  0.5318328
##   0.8805657  0.5309604
##   0.8807733  0.5302467
##   0.8711070  0.5568798
##   0.8824534  0.5296152
##   0.8886694  0.5125952
##   0.8917135  0.5040276
##   0.8939041  0.4972929
##   0.8950936  0.4932631
##   0.8959021  0.4903508
##   0.8966152  0.4877920
##   0.8972384  0.4850867
##   0.8978938  0.4826736
##   0.8979650  0.4818047
##   0.8981830  0.4807981
##   0.8982404  0.4802128
##   0.8982209  0.4794545
##   0.8980725  0.4798485
##   0.8980885  0.4789424
##   0.8980518  0.4789922
##   0.8981269  0.4785000
##   0.8981662  0.4781090
##   0.8745971  0.5515203
##   0.8854389  0.5252408
##   0.8907846  0.5107913
##   0.8938828  0.5024184
##   0.8955418  0.4960965
##   0.8968340  0.4918091
##   0.8975731  0.4884423
##   0.8981576  0.4856624
##   0.8984844  0.4837991
##   0.8987663  0.4823505
##   0.8987393  0.4814372
##   0.8987830  0.4806568
##   0.8989197  0.4796681
##   0.8988581  0.4788312
##   0.8987583  0.4787829
##   0.8987498  0.4783089
##   0.8989499  0.4776944
##   0.8988777  0.4776013
##   0.8987292  0.4774677
##   0.8878022  0.5205973
##   0.8940311  0.5020332
##   0.8967428  0.4921139
##   0.8981752  0.4862447
##   0.8989010  0.4828178
##   0.8991150  0.4811335
##   0.8995876  0.4787855
##   0.8996681  0.4778793
##   0.8996797  0.4769113
##   0.8998536  0.4757932
##   0.8998420  0.4753984
##   0.8998359  0.4750487
##   0.8997406  0.4747315
##   0.8998513  0.4745297
##   0.8999370  0.4740082
##   0.8998530  0.4739936
##   0.8999547  0.4737122
##   0.8999787  0.4733474
##   0.8999471  0.4733286
##   0.8870890  0.5193955
##   0.8928383  0.5016136
##   0.8956992  0.4920914
##   0.8969399  0.4872850
##   0.8979142  0.4835744
##   0.8980057  0.4817924
##   0.8983509  0.4804514
##   0.8986648  0.4790637
##   0.8988655  0.4777159
##   0.8989191  0.4770347
##   0.8988818  0.4769026
##   0.8988743  0.4762773
##   0.8988195  0.4759575
##   0.8988376  0.4758142
##   0.8986611  0.4757422
##   0.8985408  0.4758602
##   0.8985567  0.4755592
##   0.8985588  0.4754635
##   0.8985587  0.4756183
##   0.8933096  0.5032226
##   0.8971956  0.4895012
##   0.8989574  0.4828154
##   0.8997368  0.4787947
##   0.9002560  0.4760427
##   0.9003340  0.4748164
##   0.9004461  0.4742660
##   0.9004740  0.4732333
##   0.9005717  0.4724358
##   0.9004038  0.4723173
##   0.9004241  0.4720100
##   0.9001816  0.4723996
##   0.9000949  0.4723356
##   0.9000643  0.4723194
##   0.8999173  0.4724776
##   0.8998903  0.4723420
##   0.8997061  0.4726657
##   0.8996313  0.4729925
##   0.8996910  0.4727051
##   0.8926196  0.5033662
##   0.8964028  0.4907205
##   0.8981259  0.4834663
##   0.8986962  0.4800320
##   0.8989421  0.4775991
##   0.8989819  0.4762630
##   0.8989371  0.4754028
##   0.8989861  0.4743087
##   0.8990170  0.4739242
##   0.8988221  0.4737826
##   0.8989526  0.4731610
##   0.8989605  0.4729171
##   0.8988443  0.4730816
##   0.8987253  0.4732903
##   0.8986022  0.4733456
##   0.8986104  0.4732492
##   0.8985739  0.4730143
##   0.8986723  0.4730589
##   0.8985787  0.4731204
##
## RMSE was used to select the optimal model using the smallest value.
## The final values used for the model were n.trees = 500,
##  interaction.depth = 7, shrinkage = 0.1 and n.minobsinnode = 1.
```

Cubist

```r
library(Cubist)
# cubistMod <- cubist(solTrainXtrans, solTrainY)
# predict(cubistMod, solTestXtrans)
cubistTuned <- train(solTrainXtrans, solTrainY, method = "cubist")
cubistTuned
```

```
## Cubist
##
## 951 samples
## 228 predictors
##
## No pre-processing
## Resampling: Bootstrapped (25 reps)
## Summary of sample sizes: 951, 951, 951, 951, 951, 951, ...
## Resampling results across tuning parameters:
##
##   committees  neighbors  RMSE       Rsquared   MAE
##    1          0          0.9770705  0.7852990  0.6397877
##    1          5          0.9623528  0.7926451  0.6213113
##    1          9          0.9609770  0.7928006  0.6204098
##   10          0          0.6572851  0.8969599  0.4733539
##   10          5          0.6477720  0.8999995  0.4641000
##   10          9          0.6461732  0.9004672  0.4627475
##   20          0          0.6388975  0.9027212  0.4604583
##   20          5          0.6278051  0.9060159  0.4507676
##   20          9          0.6264733  0.9064001  0.4492627
##
## RMSE was used to select the optimal model using the smallest value.
## The final values used for the model were committees = 20 and neighbors = 9.
```

Chapter 9. A Summary of Solubility Models