Chapter 6

Chapter 6. Linear Regression and Its Cousins

yi = b0 + b1xi1 + b2xi2 + ... + bpxip + ei

Linear in the parameters: ordinary linear regression, partial least squares (pls), penalized models (ridge regression, the lasso, the elastic net)

Highly interpretable

Compute standard errors of the coefficients (make certain assumptions about the distributions of the model residuals. Can be used to assess the statistical significance of each predictor in the model)

Linear models are appropriate when the relationship between the predictors and response falls along a hyperplane

6.1 Case study: quantitative structure-activity relationship modeling

Predicting solubility using chemical structures

6.2 Linear regression

Objective of ordinary least squares linear regression: to find the plane that minimizes the sum-of-squared errors (SSE)

These estimates minimize the bias component of the bias-variance trade-off

Problems: colinearity and p>n, nonlinearity, outliers

There are no tuning parameters for multiple linear regression (validation tools are still needed)

Linear regression for solubility data

6.3 Partial least squares

Pre-processing predictors via PCA prior to performing regression is known as principal component regression (PCR) - two-step regression (dimension reduction, then regression)

PCA does not consider any aspects of the response when it selects its components. Instead, it simply chases the variability present throughout the predictor space

PLS is recommended when there are correlated predictors and a linear regression type solution is desired

like PCA, PLS finds linear combinations of the predictors. These linear combinations are commonly called components or latent variables. While the PCA linear combinations are chosen to maximally summarize predictor space variability, the PLS linear combinations of predictors are chosen to maximally summarize covariance with the response. This means that PLS finds components that maximally summarize the variation of the predictors while simultaneously requiring these components to have maximum correlation with the response.

PLS can be viewed as a supervised dimension reduction procedute, PCR is an unsupervised procedure

Prior to performing PLS, the predictors should be centered and scaled

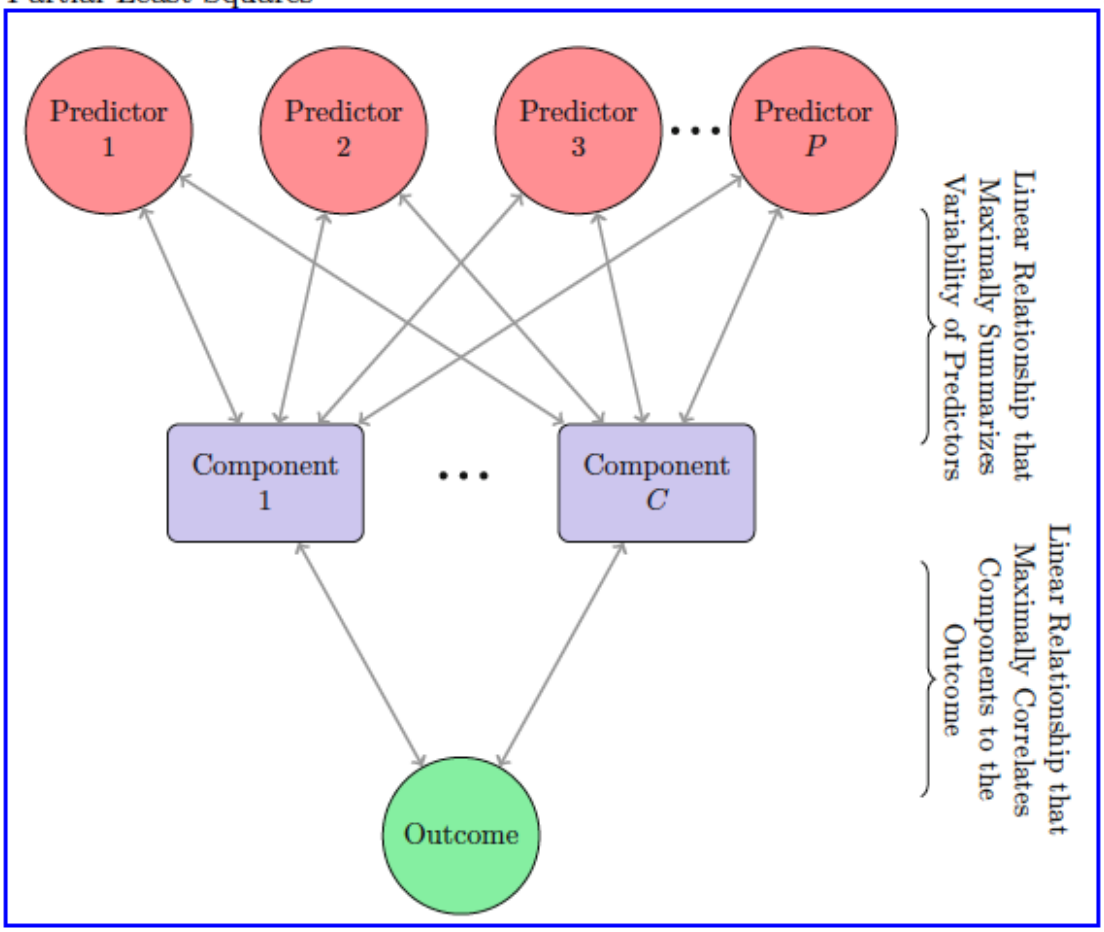


Fig. 6.9: A diagram depicting the structure of a PLS model. PLS finds components that simultaneously summarize variation of the predictors while being optimally correlated with the outcome

Figure 1: 1

PCR and PLSR for solubility data

Algorithmic variations of pls

6.4 Penalized models

Using cross-validation the penalty value is optimized

Ridge regression is known to shrink the coefficients of correlated predictors towards each other, allowing them to borrow strength from each other. In the extreme case of k identical predictors, they each get identical coefficients with 1/kth the size that any single one would get if fit alone. Lasso, on the other hand, is somewhat indifferent to very correlated predictors, and will tend to pick one and ignore the rest

Elastic net enables effective regularization via the ridge-type penalty with the feature selection quality of the lasso penalty (more effectively deal with groups of high correlated predictors)

6.5 Computing

```
library(AppliedPredictiveModeling)
data(solubility)
ls(pattern = "^solT")
```

```
## [1] "solTestX"       "solTestXtrans"  "solTestY"       "solTrainX"
## [5] "solTrainXtrans" "solTrainY"
```

```
set.seed(2)
sample(names(solTrainX),8)
```

```
## [1] "FP043"        "FP160"        "FP130"        "FP038"
## [5] "NumBonds"     "NumNonHAtoms" "FP029"        "FP185"
```

```
View(solTestX)
```

```
## Warning in system2("/usr/bin/otool", c("-L", shQuote(DSO)), stdout = TRUE):
## running command ''/usr/bin/otool' -L '/Library/Frameworks/R.framework/
## Resources/modules/R_de.so'' had status 1
```

Ordinary linear regression

```
trainingData <- solTrainXtrans
trainingData$Solubility <- solTrainY
```

```
lmFitAllPredictors <- lm(Solubility ~., data=trainingData)
summary(lmFitAllPredictors)
```

```
##
## Call:
## lm(formula = Solubility ~ ., data = trainingData)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.75620 -0.28304  0.01165  0.30030  1.54887
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)        2.431e+00  2.162e+00   1.124 0.261303
## FP001              3.594e-01  3.185e-01   1.128 0.259635
## FP002              1.456e-01  2.637e-01   0.552 0.580960
## FP003             -3.969e-02  1.314e-01  -0.302 0.762617
## FP004             -3.049e-01  1.371e-01  -2.223 0.026520 *
## FP005              2.837e+00  9.598e-01   2.956 0.003223 **
## FP006             -6.886e-02  2.041e-01  -0.337 0.735917
## FP007              4.044e-02  1.152e-01   0.351 0.725643
## FP008              1.121e-01  1.636e-01   0.685 0.493331
## FP009             -8.242e-01  8.395e-01  -0.982 0.326536
## FP010              4.193e-01  3.136e-01   1.337 0.181579
## FP011              5.158e-02  2.198e-01   0.235 0.814503
## FP012             -1.346e-02  1.611e-01  -0.084 0.933452
## FP013             -4.519e-01  5.473e-01  -0.826 0.409311
## FP014              3.281e-01  4.550e-01   0.721 0.471044
## FP015             -1.839e-01  1.521e-01  -1.209 0.226971
## FP016             -1.367e-01  1.548e-01  -0.883 0.377340
## FP017             -1.704e-01  1.386e-01  -1.230 0.219187
## FP018             -3.824e-01  2.388e-01  -1.602 0.109655
## FP019             -3.131e-01  3.863e-01  -0.811 0.417862
## FP020              2.072e-01  2.135e-01   0.971 0.332078
## FP021             -5.956e-02  2.632e-01  -0.226 0.821060
## FP022              2.336e-01  3.456e-01   0.676 0.499180
## FP023             -3.193e-01  1.909e-01  -1.672 0.094866 .
## FP024             -4.272e-01  2.827e-01  -1.511 0.131162
## FP025              4.376e-01  4.538e-01   0.964 0.335184
## FP026              2.068e-01  2.564e-01   0.806 0.420273
## FP027              2.424e-01  2.429e-01   0.998 0.318594
## FP028              1.070e-01  1.200e-01   0.892 0.372547
## FP029             -9.857e-02  2.199e-01  -0.448 0.654163
## FP030             -2.361e-01  2.468e-01  -0.957 0.339048
## FP031              8.690e-02  1.346e-01   0.646 0.518754
## FP032             -1.204e+00  7.772e-01  -1.550 0.121628
## FP033              5.766e-01  4.236e-01   1.361 0.173882
## FP034             -1.794e-01  2.618e-01  -0.685 0.493486
## FP035             -2.140e-01  1.704e-01  -1.256 0.209605
## FP036              7.701e-02  1.657e-01   0.465 0.642133
## FP037              1.098e-01  1.725e-01   0.636 0.524693
## FP038              2.721e-01  1.888e-01   1.441 0.150030
## FP039              2.011e-02  2.888e-01   0.070 0.944491
## FP040              5.477e-01  1.890e-01   2.898 0.003873 **
## FP041             -4.265e-01  3.004e-01  -1.420 0.156143
## FP042             -9.901e-01  7.078e-01  -1.399 0.162294
## FP043             -3.725e-02  2.096e-01  -0.178 0.859011
## FP044             -3.860e-01  2.184e-01  -1.768 0.077562 .
## FP045              2.120e-01  1.299e-01   1.631 0.103238
## FP046             -3.504e-02  2.733e-01  -0.128 0.898010
```

```
## FP047            -1.675e-02  1.414e-01  -0.118 0.905775
## FP048             2.610e-01  2.434e-01   1.073 0.283810
## FP049             1.241e-01  1.971e-01   0.630 0.529036
## FP050             9.087e-03  1.410e-01   0.064 0.948648
## FP051             1.050e-01  2.014e-01   0.521 0.602210
## FP052            -4.569e-01  2.482e-01  -1.841 0.066029 .
## FP053             2.994e-01  2.466e-01   1.214 0.225129
## FP054             2.734e-02  1.829e-01   0.149 0.881229
## FP055            -3.662e-01  1.970e-01  -1.858 0.063530 .
## FP056            -2.961e-01  2.979e-01  -0.994 0.320541
## FP057            -1.002e-01  1.379e-01  -0.727 0.467703
## FP058             3.100e-01  8.074e-01   0.384 0.701129
## FP059            -1.615e-01  1.690e-01  -0.956 0.339514
## FP060             2.350e-01  1.474e-01   1.595 0.111209
## FP061            -6.365e-01  1.440e-01  -4.421 1.13e-05 ***
## FP062            -5.224e-01  2.961e-01  -1.764 0.078078 .
## FP063            -2.001e+00  1.287e+00  -1.554 0.120553
## FP064             2.549e-01  1.221e-01   2.087 0.037207 *
## FP065            -2.844e-01  1.197e-01  -2.377 0.017714 *
## FP066             2.093e-01  1.264e-01   1.655 0.098301 .
## FP067            -1.406e-01  1.540e-01  -0.913 0.361631
## FP068             4.964e-01  2.028e-01   2.447 0.014630 *
## FP069             1.324e-01  8.824e-02   1.501 0.133885
## FP070             3.453e-03  8.088e-02   0.043 0.965963
## FP071             1.474e-01  1.237e-01   1.192 0.233775
## FP072            -9.773e-01  2.763e-01  -3.537 0.000431 ***
## FP073            -4.671e-01  2.072e-01  -2.254 0.024474 *
## FP074             1.793e-01  1.206e-01   1.487 0.137566
## FP075             1.231e-01  1.035e-01   1.188 0.235034
## FP076             5.166e-01  1.704e-01   3.031 0.002525 **
## FP077             1.644e-01  1.236e-01   1.331 0.183739
## FP078            -3.715e-01  1.588e-01  -2.339 0.019608 *
## FP079             4.254e-01  1.881e-01   2.262 0.023992 *
## FP080             3.101e-01  1.554e-01   1.996 0.046340 *
## FP081            -3.208e-01  1.117e-01  -2.873 0.004192 **
## FP082             1.243e-01  9.524e-02   1.305 0.192379
## FP083            -6.916e-01  2.134e-01  -3.241 0.001248 **
## FP084             3.626e-01  2.381e-01   1.523 0.128171
## FP085            -3.310e-01  1.428e-01  -2.317 0.020785 *
## FP086             1.169e-02  9.774e-02   0.120 0.904834
## FP087             4.559e-02  2.797e-01   0.163 0.870568
## FP088             2.416e-01  9.959e-02   2.425 0.015534 *
## FP089             5.999e-01  2.320e-01   2.586 0.009915 **
## FP090            -2.450e-02  1.154e-01  -0.212 0.831930
## FP091            -2.858e-01  3.185e-01  -0.897 0.369847
## FP092             2.665e-01  2.069e-01   1.288 0.198156
## FP093             1.974e-01  1.087e-01   1.816 0.069803 .
## FP094            -1.991e-01  1.441e-01  -1.381 0.167707
## FP095            -1.403e-01  1.124e-01  -1.248 0.212449
## FP096            -5.024e-01  1.459e-01  -3.445 0.000605 ***
## FP097            -2.635e-01  1.666e-01  -1.582 0.114020
## FP098            -2.865e-01  1.633e-01  -1.754 0.079863 .
## FP099             2.592e-01  2.568e-01   1.009 0.313136
## FP100            -4.008e-01  3.034e-01  -1.321 0.186949
```

```
## FP101            -1.760e-01  3.019e-01  -0.583 0.560147
## FP102             2.445e-01  3.449e-01   0.709 0.478579
## FP103            -1.493e-01  9.148e-02  -1.632 0.103176
## FP104            -1.428e-01  1.176e-01  -1.214 0.225238
## FP105            -6.912e-02  1.395e-01  -0.495 0.620482
## FP106             1.128e-01  1.288e-01   0.876 0.381495
## FP107             2.778e+00  8.247e-01   3.369 0.000796 ***
## FP108             8.836e-03  1.852e-01   0.048 0.961970
## FP109             8.200e-01  2.267e-01   3.617 0.000319 ***
## FP110             3.680e-01  3.311e-01   1.111 0.266811
## FP111            -5.565e-01  1.420e-01  -3.918 9.80e-05 ***
## FP112            -1.079e-01  2.705e-01  -0.399 0.690108
## FP113             1.511e-01  9.481e-02   1.594 0.111478
## FP114            -1.201e-01  1.891e-01  -0.635 0.525628
## FP115            -1.896e-01  1.405e-01  -1.349 0.177736
## FP116             7.778e-03  1.897e-01   0.041 0.967300
## FP117             2.583e-01  1.779e-01   1.452 0.147070
## FP118            -1.964e-01  1.230e-01  -1.596 0.110940
## FP119             7.515e-01  2.630e-01   2.857 0.004402 **
## FP120            -1.814e-01  1.794e-01  -1.011 0.312362
## FP121            -4.731e-02  3.957e-01  -0.120 0.904866
## FP122             1.048e-01  1.041e-01   1.007 0.314268
## FP123             3.926e-02  1.765e-01   0.222 0.824066
## FP124             1.235e-01  1.705e-01   0.724 0.469243
## FP125            -2.633e-04  1.151e-01  -0.002 0.998175
## FP126            -2.782e-01  1.177e-01  -2.363 0.018373 *
## FP127            -6.123e-01  1.739e-01  -3.521 0.000457 ***
## FP128            -5.424e-01  1.932e-01  -2.807 0.005136 **
## FP129            -6.731e-02  2.243e-01  -0.300 0.764167
## FP130            -1.034e+00  4.106e-01  -2.518 0.012009 *
## FP131             2.158e-01  1.617e-01   1.335 0.182405
## FP132            -1.976e-01  2.382e-01  -0.830 0.406998
## FP133            -1.573e-01  1.217e-01  -1.293 0.196319
## FP134             2.496e+00  1.196e+00   2.086 0.037310 *
## FP135             1.818e-01  1.319e-01   1.379 0.168460
## FP136            -7.763e-02  3.131e-01  -0.248 0.804237
## FP137            -4.613e-02  2.978e-01  -0.155 0.876947
## FP138            -9.392e-02  1.906e-01  -0.493 0.622251
## FP139             7.659e-02  4.063e-01   0.189 0.850517
## FP140             3.145e-01  2.149e-01   1.463 0.143784
## FP141             2.219e-01  2.765e-01   0.802 0.422532
## FP142             6.272e-01  1.488e-01   4.214 2.83e-05 ***
## FP143             9.981e-01  2.929e-01   3.407 0.000692 ***
## FP144             2.207e-01  2.839e-01   0.777 0.437195
## FP145            -1.146e-01  1.188e-01  -0.964 0.335169
## FP146            -2.324e-01  2.086e-01  -1.114 0.265716
## FP147             1.502e-01  1.228e-01   1.223 0.221703
## FP148            -1.600e-01  1.319e-01  -1.213 0.225560
## FP149             1.172e-01  1.650e-01   0.710 0.477770
## FP150             9.046e-02  1.577e-01   0.574 0.566368
## FP151             2.899e-01  3.120e-01   0.929 0.353202
## FP152            -2.544e-01  2.990e-01  -0.851 0.395087
## FP153            -3.765e-01  2.773e-01  -1.358 0.175029
## FP154            -1.027e+00  2.033e-01  -5.054 5.50e-07 ***
```

```
## FP155            4.888e-01  2.916e-01   1.676 0.094163 .
## FP156           -3.602e-02  3.636e-01  -0.099 0.921109
## FP157           -4.715e-01  2.468e-01  -1.910 0.056505 .
## FP158            1.669e-02  1.925e-01   0.087 0.930943
## FP159            1.800e-01  2.432e-01   0.740 0.459378
## FP160            1.525e-02  2.177e-01   0.070 0.944155
## FP161           -2.440e-01  1.433e-01  -1.703 0.089063 .
## FP162            4.910e-02  1.859e-01   0.264 0.791710
## FP163            4.785e-01  3.121e-01   1.533 0.125659
## FP164            5.096e-01  1.899e-01   2.684 0.007446 **
## FP165            5.793e-01  2.146e-01   2.700 0.007103 **
## FP166           -6.582e-02  2.185e-01  -0.301 0.763293
## FP167           -6.044e-01  2.515e-01  -2.403 0.016502 *
## FP168           -1.187e-01  1.872e-01  -0.634 0.526173
## FP169           -1.705e-01  8.312e-02  -2.051 0.040650 *
## FP170           -7.902e-02  1.560e-01  -0.506 0.612745
## FP171            4.651e-01  1.186e-01   3.922 9.64e-05 ***
## FP172           -4.426e-01  2.440e-01  -1.814 0.070120 .
## FP173            4.243e-01  1.657e-01   2.561 0.010634 *
## FP174           -1.010e-01  2.098e-01  -0.481 0.630311
## FP175           -4.657e-02  2.481e-01  -0.188 0.851136
## FP176            9.736e-01  2.644e-01   3.682 0.000249 ***
## FP177            1.386e-01  2.393e-01   0.579 0.562538
## FP178            6.497e-02  2.079e-01   0.313 0.754691
## FP179           -3.415e-02  2.232e-01  -0.153 0.878437
## FP180           -7.905e-01  5.523e-01  -1.431 0.152839
## FP181            4.925e-01  3.218e-01   1.531 0.126309
## FP182           -1.124e-01  1.310e-01  -0.858 0.391384
## FP183            2.998e-01  7.143e-01   0.420 0.674836
## FP184            4.876e-01  1.580e-01   3.087 0.002103 **
## FP185           -3.778e-01  2.037e-01  -1.854 0.064108 .
## FP186           -3.654e-01  1.953e-01  -1.871 0.061710 .
## FP187            4.457e-01  2.682e-01   1.662 0.097015 .
## FP188            1.475e-01  1.258e-01   1.172 0.241519
## FP189           -1.984e-02  3.468e-01  -0.057 0.954384
## FP190            2.629e-01  3.018e-01   0.871 0.383981
## FP191            2.799e-01  1.465e-01   1.911 0.056388 .
## FP192           -2.404e-01  2.751e-01  -0.874 0.382534
## FP193            1.502e-01  1.494e-01   1.005 0.315159
## FP194            8.029e-01  6.379e-01   1.259 0.208566
## FP195            5.967e-02  3.435e-01   0.174 0.862158
## FP196            1.091e-02  2.544e-01   0.043 0.965812
## FP197           -3.736e-02  1.569e-01  -0.238 0.811793
## FP198            1.896e-01  2.665e-01   0.712 0.476893
## FP199           -9.932e-02  1.797e-01  -0.553 0.580702
## FP200           -6.421e-02  2.161e-01  -0.297 0.766462
## FP201           -4.838e-01  1.980e-01  -2.444 0.014771 *
## FP202            5.664e-01  1.869e-01   3.031 0.002527 **
## FP203            2.586e-01  6.447e-01   0.401 0.688462
## FP204           -1.371e-01  2.543e-01  -0.539 0.590008
## FP205            7.177e-02  1.561e-01   0.460 0.645857
## FP206           -6.769e-02  1.860e-01  -0.364 0.716094
## FP207           -5.538e-03  2.060e-01  -0.027 0.978560
## FP208           -5.338e-01  6.324e-01  -0.844 0.398925
```

```
## MolWeight         -1.232e+00  2.296e-01  -5.365 1.09e-07 ***
## NumAtoms          -1.478e+01  3.473e+00  -4.257 2.35e-05 ***
## NumNonHAtoms       1.795e+01  3.166e+00   5.670 2.07e-08 ***
## NumBonds           9.843e+00  2.681e+00   3.671 0.000260 ***
## NumNonHBonds      -1.030e+01  1.793e+00  -5.746 1.35e-08 ***
## NumMultBonds       2.107e-01  1.754e-01   1.201 0.229990
## NumRotBonds       -5.213e-01  1.334e-01  -3.908 0.000102 ***
## NumDblBonds       -7.492e-01  3.163e-01  -2.369 0.018111 *
## NumAromaticBonds  -2.364e+00  6.232e-01  -3.794 0.000161 ***
## NumHydrogen        8.347e-01  1.880e-01   4.439 1.04e-05 ***
## NumCarbon          1.730e-02  3.763e-01   0.046 0.963335
## NumNitrogen        6.125e+00  3.045e+00   2.011 0.044645 *
## NumOxygen          2.389e+00  4.523e-01   5.283 1.69e-07 ***
## NumSulfer         -8.508e+00  3.619e+00  -2.351 0.018994 *
## NumChlorine       -7.449e+00  1.989e+00  -3.744 0.000195 ***
## NumHalogen         1.408e+00  2.109e+00   0.668 0.504615
## NumRings           1.276e+00  6.716e-01   1.901 0.057731 .
## HydrophilicFactor  1.099e-02  1.137e-01   0.097 0.922998
## SurfaceArea1       8.825e-02  6.058e-02   1.457 0.145643
## SurfaceArea2       9.555e-02  5.615e-02   1.702 0.089208 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5524 on 722 degrees of freedom
## Multiple R-squared:  0.9446, Adjusted R-squared:  0.9271
## F-statistic: 54.03 on 228 and 722 DF,  p-value: < 2.2e-16
```

```
lmPred1 <- predict(lmFitAllPredictors, solTestXtrans)
head(lmPred1)
```

```
##          20          21          23          25          28          31
##  0.99370933  0.06834627 -0.69877632  0.84796356 -0.16578324  1.40815083
```

```
library(caret)
```

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

```
## Loading required package: ggplot2
```

```
lmValues1 <- data.frame(obs = solTestY, pred = lmPred1)
defaultSummary(lmValues1)
```

```
##      RMSE  Rsquared       MAE
## 0.7455802 0.8722236 0.5497605
```

```
library(MASS)
rlmFitAllPredictors <- rlm(Solubility ~., data = trainingData)
summary(rlmFitAllPredictors)
```

```
##
## Call: rlm(formula = Solubility ~ ., data = trainingData)
```

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.89940 -0.25046  0.01221  0.25351  1.86225
##
## Coefficients:
##                   Value    Std. Error t value
## (Intercept)         2.5861   1.9646     1.3164
## FP001               0.3706   0.2894     1.2804
## FP002               0.0370   0.2396     0.1546
## FP003              -0.0527   0.1194    -0.4419
## FP004              -0.2927   0.1246    -2.3491
## FP005               2.2348   0.8721     2.5626
## FP006              -0.1329   0.1854    -0.7167
## FP007               0.0144   0.1047     0.1376
## FP008               0.1517   0.1486     1.0209
## FP009              -0.8072   0.7628    -1.0582
## FP010               0.2696   0.2849     0.9464
## FP011               0.2455   0.1997     1.2294
## FP012              -0.0579   0.1464    -0.3957
## FP013              -0.2125   0.4973    -0.4272
## FP014               0.2084   0.4134     0.5041
## FP015              -0.2071   0.1382    -1.4984
## FP016              -0.2203   0.1406    -1.5670
## FP017              -0.1594   0.1259    -1.2662
## FP018              -0.4960   0.2169    -2.2862
## FP019              -0.7774   0.3510    -2.2150
## FP020               0.0829   0.1939     0.4275
## FP021               0.0499   0.2392     0.2086
## FP022               0.3125   0.3140     0.9954
## FP023              -0.3382   0.1735    -1.9496
## FP024              -0.1680   0.2569    -0.6540
## FP025               0.1863   0.4123     0.4517
## FP026               0.3676   0.2330     1.5777
## FP027               0.3448   0.2207     1.5626
## FP028               0.0704   0.1090     0.6454
## FP029               0.1165   0.1998     0.5830
## FP030              -0.2305   0.2243    -1.0279
## FP031               0.1540   0.1223     1.2595
## FP032              -1.2437   0.7062    -1.7612
## FP033               0.5611   0.3849     1.4577
## FP034              -0.2641   0.2379    -1.1100
## FP035              -0.2015   0.1548    -1.3018
## FP036               0.2637   0.1505     1.7522
## FP037               0.1458   0.1567     0.9301
## FP038               0.4082   0.1716     2.3793
## FP039               0.1305   0.2624     0.4974
## FP040               0.4512   0.1717     2.6269
## FP041              -0.3120   0.2730    -1.1430
## FP042              -0.9665   0.6431    -1.5028
## FP043              -0.2024   0.1905    -1.0626
## FP044              -0.4009   0.1984    -2.0206
## FP045               0.2682   0.1180     2.2721
## FP046              -0.1660   0.2483    -0.6683
## FP047              -0.0429   0.1285    -0.3340
```

```
## FP048              0.2185   0.2211    0.9880
## FP049              0.2413   0.1791    1.3471
## FP050             -0.0402   0.1281   -0.3139
## FP051              0.1457   0.1830    0.7965
## FP052             -0.4612   0.2255   -2.0450
## FP053              0.2263   0.2241    1.0099
## FP054              0.0267   0.1662    0.1607
## FP055             -0.3384   0.1790   -1.8900
## FP056             -0.5154   0.2707   -1.9039
## FP057             -0.1288   0.1253   -1.0275
## FP058              0.3464   0.7336    0.4722
## FP059             -0.1004   0.1535   -0.6541
## FP060              0.1954   0.1339    1.4589
## FP061             -0.7038   0.1308   -5.3797
## FP062             -0.4596   0.2690   -1.7084
## FP063             -3.0917   1.1697   -2.6431
## FP064              0.1505   0.1110    1.3566
## FP065             -0.2066   0.1087   -1.9002
## FP066              0.2428   0.1149    2.1137
## FP067             -0.2495   0.1400   -1.7829
## FP068              0.6864   0.1843    3.7246
## FP069              0.1494   0.0802    1.8637
## FP070             -0.0558   0.0735   -0.7590
## FP071              0.1477   0.1124    1.3145
## FP072             -1.1837   0.2511   -4.7142
## FP073             -0.5319   0.1883   -2.8252
## FP074              0.2918   0.1096    2.6629
## FP075              0.0938   0.0941    0.9967
## FP076              0.5134   0.1549    3.3151
## FP077              0.2343   0.1123    2.0869
## FP078             -0.4152   0.1443   -2.8775
## FP079              0.3700   0.1709    2.1649
## FP080              0.3392   0.1412    2.4030
## FP081             -0.2443   0.1015   -2.4076
## FP082              0.0836   0.0865    0.9655
## FP083             -0.6630   0.1939   -3.4190
## FP084              0.2897   0.2163    1.3390
## FP085             -0.3261   0.1298   -2.5124
## FP086              0.0042   0.0888    0.0473
## FP087              0.1507   0.2542    0.5930
## FP088              0.2261   0.0905    2.4989
## FP089              0.5282   0.2108    2.5052
## FP090             -0.0621   0.1049   -0.5922
## FP091             -0.4952   0.2894   -1.7115
## FP092              0.2044   0.1880    1.0873
## FP093              0.1163   0.0988    1.1777
## FP094             -0.1073   0.1310   -0.8190
## FP095             -0.0888   0.1021   -0.8689
## FP096             -0.5609   0.1325   -4.2319
## FP097             -0.2391   0.1513   -1.5799
## FP098             -0.3220   0.1484   -2.1700
## FP099              0.5687   0.2333    2.4376
## FP100             -0.2545   0.2757   -0.9231
## FP101              0.0425   0.2743    0.1550
```

```
## FP102              0.2444   0.3134    0.7798
## FP103             -0.1706   0.0831   -2.0520
## FP104             -0.2192   0.1069   -2.0514
## FP105              0.0315   0.1268    0.2485
## FP106              0.1044   0.1170    0.8921
## FP107              2.4059   0.7493    3.2108
## FP108              0.0581   0.1683    0.3455
## FP109              0.9248   0.2060    4.4899
## FP110              0.2497   0.3009    0.8301
## FP111             -0.4888   0.1291   -3.7873
## FP112             -0.2874   0.2458   -1.1692
## FP113              0.1009   0.0861    1.1717
## FP114             -0.2667   0.1718   -1.5522
## FP115             -0.2004   0.1277   -1.5696
## FP116              0.1425   0.1723    0.8268
## FP117              0.3100   0.1617    1.9176
## FP118             -0.1368   0.1118   -1.2239
## FP119              0.5325   0.2390    2.2280
## FP120             -0.1572   0.1630   -0.9644
## FP121             -0.0857   0.3596   -0.2384
## FP122              0.1049   0.0946    1.1090
## FP123             -0.0723   0.1604   -0.4505
## FP124              0.1504   0.1549    0.9709
## FP125             -0.0208   0.1046   -0.1990
## FP126             -0.3416   0.1069   -3.1946
## FP127             -0.5554   0.1580   -3.5151
## FP128             -0.5344   0.1756   -3.0434
## FP129             -0.0289   0.2038   -0.1419
## FP130             -0.6492   0.3731   -1.7401
## FP131              0.2098   0.1469    1.4278
## FP132             -0.2389   0.2164   -1.1038
## FP133             -0.1433   0.1105   -1.2965
## FP134              3.0068   1.0871    2.7659
## FP135              0.0407   0.1198    0.3400
## FP136             -0.1699   0.2845   -0.5973
## FP137              0.0880   0.2706    0.3252
## FP138             -0.1248   0.1731   -0.7209
## FP139             -0.2078   0.3691   -0.5630
## FP140              0.4015   0.1953    2.0558
## FP141              0.2224   0.2513    0.8851
## FP142              0.7016   0.1352    5.1881
## FP143              1.1801   0.2661    4.4339
## FP144              0.3078   0.2579    1.1934
## FP145             -0.0268   0.1079   -0.2478
## FP146             -0.2993   0.1895   -1.5792
## FP147              0.1306   0.1116    1.1706
## FP148             -0.1155   0.1199   -0.9637
## FP149              0.0434   0.1499    0.2898
## FP150              0.1316   0.1433    0.9187
## FP151              0.3921   0.2835    1.3832
## FP152             -0.2870   0.2716   -1.0567
## FP153             -0.5698   0.2520   -2.2614
## FP154             -1.2141   0.1847   -6.5740
## FP155              0.5297   0.2650    1.9989
```

```
## FP156          -0.5901   0.3304   -1.7861
## FP157          -0.4511   0.2243   -2.0115
## FP158          -0.0963   0.1749   -0.5502
## FP159           0.0603   0.2210    0.2730
## FP160          -0.0118   0.1978   -0.0596
## FP161          -0.3881   0.1302   -2.9814
## FP162           0.0655   0.1689    0.3881
## FP163           0.3068   0.2836    1.0820
## FP164           0.6726   0.1725    3.8982
## FP165           0.5248   0.1950    2.6915
## FP166           0.0297   0.1985    0.1496
## FP167          -0.5843   0.2285   -2.5570
## FP168          -0.1659   0.1701   -0.9754
## FP169          -0.1580   0.0755   -2.0928
## FP170           0.0152   0.1418    0.1073
## FP171           0.4332   0.1078    4.0192
## FP172          -0.4863   0.2217   -2.1935
## FP173           0.4567   0.1505    3.0338
## FP174          -0.1927   0.1906   -1.0109
## FP175          -0.1397   0.2254   -0.6196
## FP176           1.1228   0.2403    4.6732
## FP177           0.0941   0.2174    0.4329
## FP178          -0.0629   0.1889   -0.3331
## FP179          -0.1814   0.2028   -0.8945
## FP180          -0.2895   0.5019   -0.5768
## FP181           0.2199   0.2924    0.7521
## FP182          -0.1557   0.1191   -1.3079
## FP183           0.7978   0.6490    1.2292
## FP184           0.4332   0.1435    3.0182
## FP185          -0.3395   0.1851   -1.8339
## FP186          -0.2692   0.1774   -1.5169
## FP187           0.0303   0.2437    0.1242
## FP188           0.0787   0.1143    0.6887
## FP189           0.0945   0.3151    0.3000
## FP190           0.3688   0.2742    1.3450
## FP191           0.3227   0.1331    2.4248
## FP192          -0.3142   0.2500   -1.2569
## FP193           0.1705   0.1358    1.2561
## FP194           0.8636   0.5796    1.4900
## FP195          -0.1132   0.3121   -0.3625
## FP196          -0.0928   0.2312   -0.4015
## FP197          -0.1103   0.1425   -0.7741
## FP198           0.1807   0.2421    0.7464
## FP199          -0.0008   0.1633   -0.0051
## FP200          -0.2167   0.1963   -1.1037
## FP201          -0.5956   0.1799   -3.3108
## FP202           0.6575   0.1698    3.8721
## FP203           0.2424   0.5858    0.4138
## FP204          -0.0565   0.2311   -0.2446
## FP205           0.1484   0.1418    1.0462
## FP206           0.0252   0.1690    0.1493
## FP207          -0.0322   0.1872   -0.1718
## FP208          -0.5715   0.5747   -0.9946
## MolWeight      -1.2955   0.2086   -6.2095
```

```
## NumAtoms          -16.8343   3.1558    -5.3344
## NumNonHAtoms       20.4017   2.8765     7.0926
## NumBonds           10.7076   2.4364     4.3948
## NumNonHBonds      -11.6342   1.6289    -7.1425
## NumMultBonds        0.0481   0.1593     0.3016
## NumRotBonds        -0.5600   0.1212    -4.6202
## NumDblBonds        -0.6851   0.2874    -2.3840
## NumAromaticBonds   -2.0220   0.5663    -3.5706
## NumHydrogen         0.7778   0.1709     4.5527
## NumCarbon           0.7865   0.3419     2.3003
## NumNitrogen         8.5838   2.7669     3.1023
## NumOxygen           2.6481   0.4110     6.4436
## NumSulfer          -9.6687   3.2884    -2.9403
## NumChlorine        -6.4608   1.8075    -3.5744
## NumHalogen          1.4341   1.9165     0.7483
## NumRings            1.0132   0.6102     1.6605
## HydrophilicFactor  -0.0836   0.1033    -0.8094
## SurfaceArea1        0.0948   0.0550     1.7225
## SurfaceArea2        0.1181   0.0510     2.3141
##
## Residual standard error: 0.3739 on 722 degrees of freedom
```

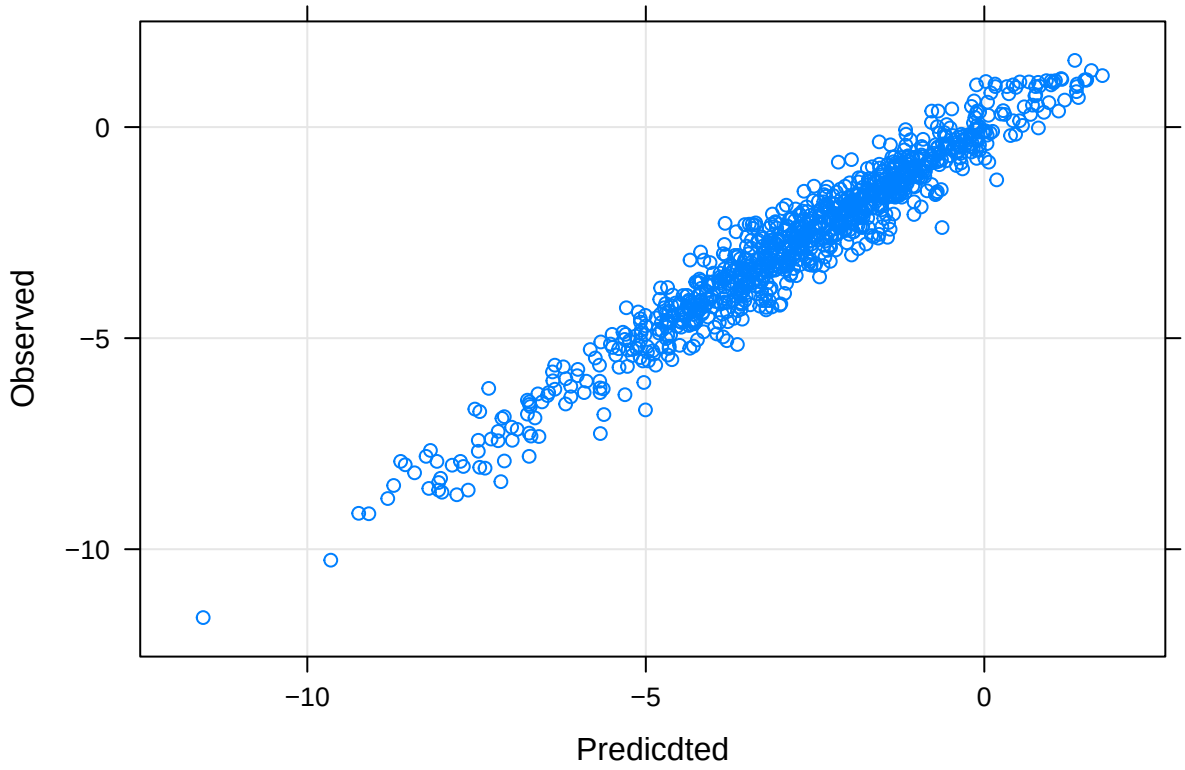
```
ctrl <- trainControl(method = "cv", number = 10)
set.seed(100)
lmFit1 <- train(x = solTrainXtrans, y = solTrainY, method = "lm", trControl = ctrl)
```

```
## 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
```

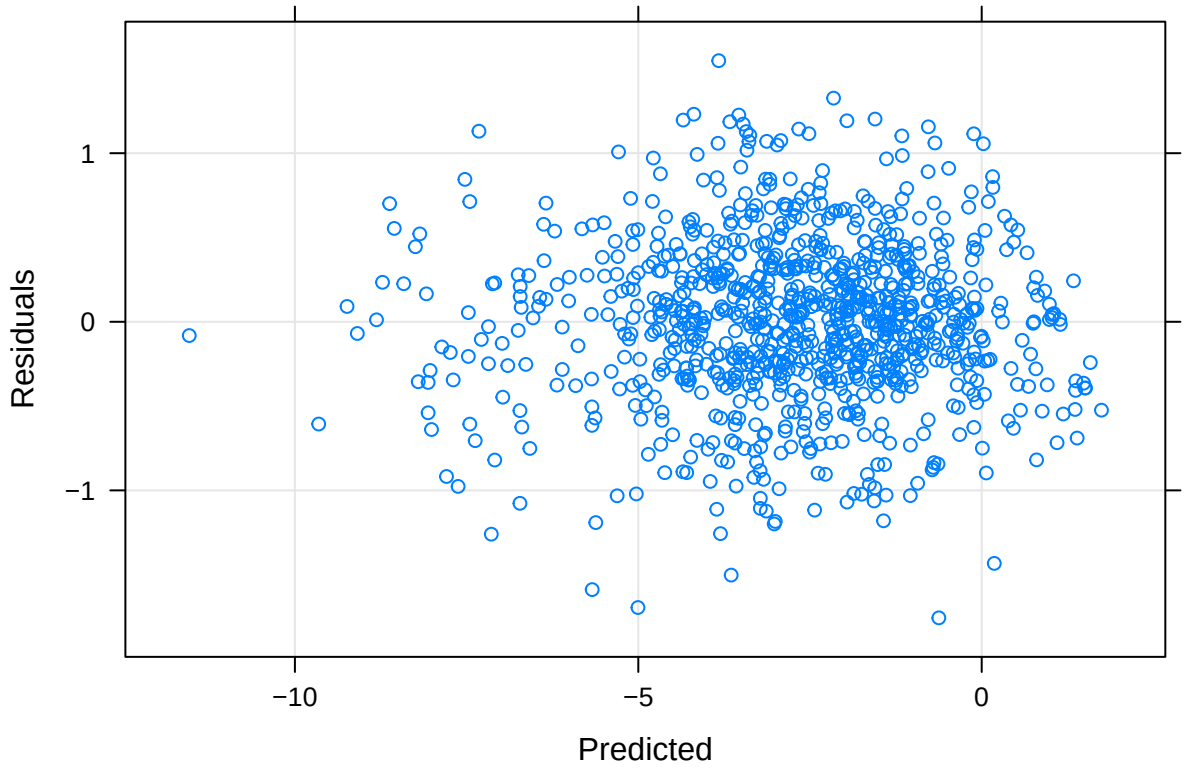
```
lmFit1
```

```
## Linear Regression
##
## 951 samples
## 228 predictors
##
## No pre-processing
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 856, 856, 855, 855, 857, 856, ...
## Resampling results:
##
##   RMSE       Rsquared   MAE
##   0.6926348  0.8872058  0.5199216
##
## Tuning parameter 'intercept' was held constant at a value of TRUE
```

```
xyplot(solTrainY ~predict(lmFit1), type = c("p","g"), xlab = "Predicdted", ylab = "Observed")
```




```
xyplot(resid(lmFit1) ~ predict(lmFit1), type = c("p","g"), xlab = "Predicted", ylab = "Residuals")
```



```
corThresh <- .9
tooHigh <- findCorrelation(cor(solTrainXtrans),corThresh)
corrPred <- names(solTrainXtrans)[tooHigh]
trainXfiltered <- solTrainXtrans[,-tooHigh]
testXfiltered <- solTestXtrans[,-tooHigh]
set.seed(100)
lmFiltered <- train(solTrainXtrans, solTrainY, method = "lm", trControl = ctrl)
```

```
## 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
```

```
lmFiltered
```

```
## Linear Regression
##
## 951 samples
## 228 predictors
##
## No pre-processing
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 856, 856, 855, 855, 857, 856, ...
## Resampling results:
##
##   RMSE       Rsquared   MAE
##   0.6926348  0.8872058  0.5199216
##
## Tuning parameter 'intercept' was held constant at a value of TRUE
```

```
set.seed(100)
rlmPCA <- train(solTrainXtrans, solTrainY, method = "rlm", preProcess = "pca", trControl = ctrl)
```

```
## Warning in rlm.default(x, y, weights, method = method, wt.method =
## wt.method, : 'rlm' failed to converge in 20 steps

## Warning in rlm.default(x, y, weights, method = method, wt.method =
## wt.method, : 'rlm' failed to converge in 20 steps
```

```
rlmPCA
```

```
## Robust Linear Model
##
## 951 samples
## 228 predictors
##
## Pre-processing: principal component signal extraction (228),
```

```
##  centered (228), scaled (228)
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 856, 856, 855, 855, 857, 856, ...
## Resampling results across tuning parameters:
##
##   intercept  psi           RMSE       Rsquared   MAE
##   FALSE      psi.huber     2.8245812  0.8561008  2.7155082
##   FALSE      psi.hampel    2.8245118  0.8561763  2.7154817
##   FALSE      psi.bisquare  2.8244621  0.8562213  2.7154216
##    TRUE      psi.huber     0.7828457  0.8550937  0.5970333
##    TRUE      psi.hampel    0.7825426  0.8552118  0.5972415
##    TRUE      psi.bisquare  0.7903976  0.8524410  0.6016888
##
## RMSE was used to select the optimal model using the smallest value.
## The final values used for the model were intercept = TRUE and psi
##  = psi.hampel.
```

Partial least squares

```
library(pls)
```

```
##
## Attaching package: 'pls'

## The following object is masked from 'package:caret':
##
##     R2

## The following object is masked from 'package:stats':
##
##     loadings
```

```
plsFit <- plsr(Solubility ~., data = trainingData)
summary(plsFit)
```

```
## Data:    X dimension: 951 228
##  Y dimension: 951 1
## Fit method: kernelpls
## Number of components considered: 228
## TRAINING: % variance explained
##             1 comps  2 comps  3 comps  4 comps  5 comps  6 comps  7 comps
## X             49.80    65.87    71.13    73.66    74.86    76.08    77.37
## Solubility    26.52    61.86    75.13    84.28    87.79    89.44    90.20
##             8 comps  9 comps  10 comps  11 comps  12 comps  13 comps
## X             78.58    80.33     81.56     82.32     82.96     83.64
## Solubility    90.81    91.17     91.52     91.97     92.34     92.56
##             14 comps  15 comps  16 comps  17 comps  18 comps  19 comps
## X              84.14     85.13     85.77     86.37     86.81     87.47
## Solubility     92.77     92.90     93.06     93.14     93.26     93.33
##             20 comps  21 comps  22 comps  23 comps  24 comps  25 comps
```

```
## X              87.78     88.28     88.63     88.89     89.14     89.51
## Solubility     93.43     93.48     93.53     93.59     93.64     93.68
##             26 comps  27 comps  28 comps  29 comps  30 comps  31 comps
## X              89.84     90.06     90.32     90.53     90.72     90.90
## Solubility     93.71     93.74     93.77     93.80     93.82     93.84
##             32 comps  33 comps  34 comps  35 comps  36 comps  37 comps
## X              91.18     91.38     91.59     91.84     92.03     92.21
## Solubility     93.86     93.87     93.89     93.90     93.91     93.92
##             38 comps  39 comps  40 comps  41 comps  42 comps  43 comps
## X              92.35     92.51     92.69     92.83     93.00     93.22
## Solubility     93.94     93.95     93.95     93.96     93.97     93.97
##             44 comps  45 comps  46 comps  47 comps  48 comps  49 comps
## X              93.38     93.52     93.69     93.85     93.97     94.12
## Solubility     93.98     93.99     93.99     94.00     94.01     94.02
##             50 comps  51 comps  52 comps  53 comps  54 comps  55 comps
## X              94.25     94.40     94.53     94.64     94.73     94.85
## Solubility     94.03     94.04     94.05     94.06     94.08     94.09
##             56 comps  57 comps  58 comps  59 comps  60 comps  61 comps
## X              94.96     95.09     95.21     95.31     95.42     95.50
## Solubility     94.10     94.11     94.12     94.13     94.14     94.15
##             62 comps  63 comps  64 comps  65 comps  66 comps  67 comps
## X              95.59     95.67     95.77     95.84     95.91     95.98
## Solubility     94.15     94.16     94.17     94.17     94.18     94.18
##             68 comps  69 comps  70 comps  71 comps  72 comps  73 comps
## X              96.06     96.12     96.21     96.28     96.34     96.43
## Solubility     94.19     94.20     94.20     94.21     94.21     94.22
##             74 comps  75 comps  76 comps  77 comps  78 comps  79 comps
## X              96.53     96.59     96.66     96.71     96.77     96.82
## Solubility     94.22     94.23     94.24     94.26     94.27     94.28
##             80 comps  81 comps  82 comps  83 comps  84 comps  85 comps
## X              96.87     96.93     96.99     97.05     97.10     97.15
## Solubility     94.30     94.31     94.32     94.33     94.34     94.35
##             86 comps  87 comps  88 comps  89 comps  90 comps  91 comps
## X              97.21     97.27     97.32     97.38     97.44     97.48
## Solubility     94.35     94.36     94.37     94.37     94.38     94.39
##             92 comps  93 comps  94 comps  95 comps  96 comps  97 comps
## X              97.52     97.56     97.61     97.66     97.71     97.76
## Solubility     94.39     94.40     94.40     94.41     94.41     94.41
##             98 comps  99 comps  100 comps  101 comps  102 comps  103 comps
## X              97.80     97.84      97.89      97.93      97.98      98.01
## Solubility     94.42     94.42      94.42      94.42      94.43      94.43
##             104 comps  105 comps  106 comps  107 comps  108 comps
## X               98.05      98.09      98.14      98.18      98.21
## Solubility      94.43      94.43      94.43      94.43      94.43
##             109 comps  110 comps  111 comps  112 comps  113 comps
## X               98.25      98.29      98.33      98.36      98.39
## Solubility      94.43      94.43      94.43      94.44      94.44
##             114 comps  115 comps  116 comps  117 comps  118 comps
## X               98.43      98.47      98.50      98.53      98.56
## Solubility      94.44      94.44      94.44      94.44      94.44
##             119 comps  120 comps  121 comps  122 comps  123 comps
## X               98.60      98.63      98.67      98.69      98.72
## Solubility      94.44      94.44      94.45      94.45      94.45
##             124 comps  125 comps  126 comps  127 comps  128 comps
```

```
## X              98.75      98.78      98.80      98.83      98.86
## Solubility     94.45      94.45      94.45      94.45      94.45
##             129 comps  130 comps  131 comps  132 comps  133 comps
## X              98.88      98.91      98.94      98.96      98.98
## Solubility     94.46      94.46      94.46      94.46      94.46
##             134 comps  135 comps  136 comps  137 comps  138 comps
## X              99.00      99.03      99.05      99.07      99.09
## Solubility     94.46      94.46      94.46      94.46      94.46
##             139 comps  140 comps  141 comps  142 comps  143 comps
## X              99.11      99.13      99.15      99.17      99.19
## Solubility     94.46      94.46      94.46      94.46      94.46
##             144 comps  145 comps  146 comps  147 comps  148 comps
## X              99.21      99.23      99.25      99.27      99.28
## Solubility     94.46      94.46      94.46      94.46      94.46
##             149 comps  150 comps  151 comps  152 comps  153 comps
## X              99.30      99.32      99.33      99.35      99.36
## Solubility     94.46      94.46      94.46      94.46      94.46
##             154 comps  155 comps  156 comps  157 comps  158 comps
## X              99.38      99.39      99.41      99.42      99.43
## Solubility     94.46      94.46      94.46      94.46      94.46
##             159 comps  160 comps  161 comps  162 comps  163 comps
## X              99.45      99.46      99.47      99.49      99.50
## Solubility     94.46      94.46      94.46      94.46      94.46
##             164 comps  165 comps  166 comps  167 comps  168 comps
## X              99.52      99.53      99.54      99.56      99.57
## Solubility     94.46      94.46      94.46      94.46      94.46
##             169 comps  170 comps  171 comps  172 comps  173 comps
## X              99.58      99.60      99.61      99.62      99.63
## Solubility     94.46      94.46      94.46      94.46      94.46
##             174 comps  175 comps  176 comps  177 comps  178 comps
## X              99.64      99.65      99.66      99.67      99.68
## Solubility     94.46      94.46      94.46      94.46      94.46
##             179 comps  180 comps  181 comps  182 comps  183 comps
## X              99.69      99.70      99.71      99.72      99.73
## Solubility     94.46      94.46      94.46      94.46      94.46
##             184 comps  185 comps  186 comps  187 comps  188 comps
## X              99.74      99.75      99.76      99.77      99.77
## Solubility     94.46      94.46      94.46      94.46      94.46
##             189 comps  190 comps  191 comps  192 comps  193 comps
## X              99.78      99.79      99.80      99.81      99.81
## Solubility     94.46      94.46      94.46      94.46      94.46
##             194 comps  195 comps  196 comps  197 comps  198 comps
## X              99.82      99.83      99.84      99.85      99.85
## Solubility     94.46      94.46      94.46      94.46      94.46
##             199 comps  200 comps  201 comps  202 comps  203 comps
## X              99.86      99.86      99.87      99.87      99.88
## Solubility     94.46      94.46      94.46      94.46      94.46
##             204 comps  205 comps  206 comps  207 comps  208 comps
## X              99.88      99.89      99.90      99.90      99.91
## Solubility     94.46      94.46      94.46      94.46      94.46
##             209 comps  210 comps  211 comps  212 comps  213 comps
## X              99.91      99.92      99.93      99.93      99.94
## Solubility     94.46      94.46      94.46      94.46      94.46
##             214 comps  215 comps  216 comps  217 comps  218 comps
```

```
## X               99.94      99.95      99.95      99.96      99.96
## Solubility      94.46      94.46      94.46      94.46      94.46
##             219 comps  220 comps  221 comps  222 comps  223 comps
## X               99.97      99.97      99.98      99.98      99.98
## Solubility      94.46      94.46      94.46      94.46      94.46
##             224 comps  225 comps  226 comps  227 comps  228 comps
## X               99.99      99.99      99.99     100.00     100.00
## Solubility      94.46      94.46      94.46      94.46      94.46
```

```
names(plsFit)
```

```
##  [1] "coefficients"    "scores"          "loadings"
##  [4] "loading.weights" "Yscores"         "Yloadings"
##  [7] "projection"      "Xmeans"          "Ymeans"
## [10] "fitted.values"   "residuals"       "Xvar"
## [13] "Xtotvar"         "fit.time"        "ncomp"
## [16] "method"          "call"            "terms"
## [19] "model"
```

```
predict(plsFit, solTestXtrans[1:5,], ncomp = 1:2)
```

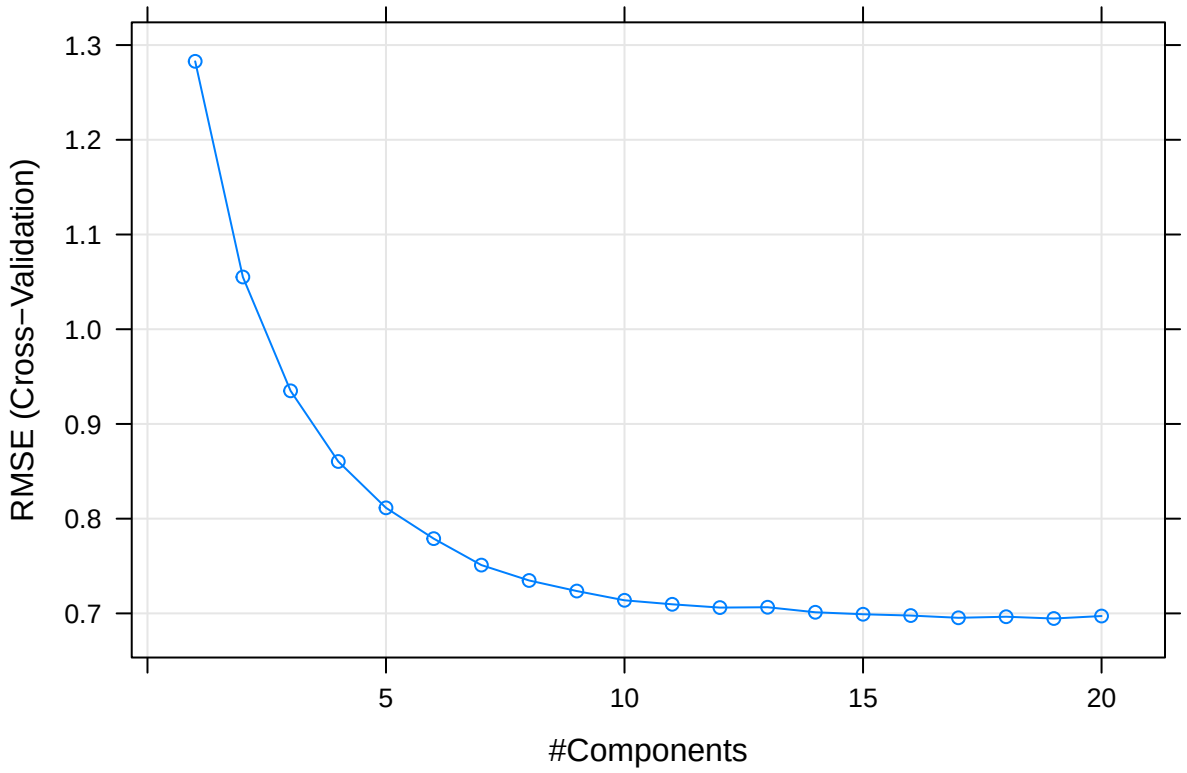
```
## , , 1 comps
##
##    Solubility
## 20  -1.789335
## 21  -1.427551
## 23  -2.268798
## 25  -2.269782
## 28  -1.867960
##
## , , 2 comps
##
##    Solubility
## 20  0.2520469
## 21  0.3555028
## 23 -1.8795338
## 25 -0.6848584
## 28 -1.5531552
```

```
set.seed(100)
plsTune <- train(solTrainXtrans, solTrainY, method = "pls", tuneLength = 20,
                 # The default tuning grid evaluates components 1 ... tuneLength)
                 trControl = ctrl, preProc = c("center","scale"))
plsTune
```

```
## Partial Least Squares
##
## 951 samples
## 228 predictors
##
## Pre-processing: centered (228), scaled (228)
## Resampling: Cross-Validated (10 fold)
```

```
## Summary of sample sizes: 856, 856, 855, 855, 857, 856, ...
## Resampling results across tuning parameters:
##
##   ncomp  RMSE       Rsquared   MAE
##    1     1.2828145  0.6079795  0.9893636
##    2     1.0551277  0.7378376  0.8297133
##    3     0.9349505  0.7939934  0.7229185
##    4     0.8603662  0.8254588  0.6695206
##    5     0.8114226  0.8443879  0.6341178
##    6     0.7789089  0.8568821  0.6043381
##    7     0.7509779  0.8674586  0.5737601
##    8     0.7347473  0.8730535  0.5616286
##    9     0.7235864  0.8772237  0.5525797
##   10     0.7138120  0.8803802  0.5489714
##   11     0.7096044  0.8818434  0.5459290
##   12     0.7061430  0.8832626  0.5419456
##   13     0.7065012  0.8838061  0.5403922
##   14     0.7011695  0.8855274  0.5360594
##   15     0.6990833  0.8859888  0.5310024
##   16     0.6977189  0.8865601  0.5326943
##   17     0.6953522  0.8874532  0.5316631
##   18     0.6964785  0.8869715  0.5329199
##   19     0.6945869  0.8874051  0.5301209
##   20     0.6972065  0.8864873  0.5320990
##
## RMSE was used to select the optimal model using the smallest value.
## The final value used for the model was ncomp = 19.
```

```
plot(plsTune)
```



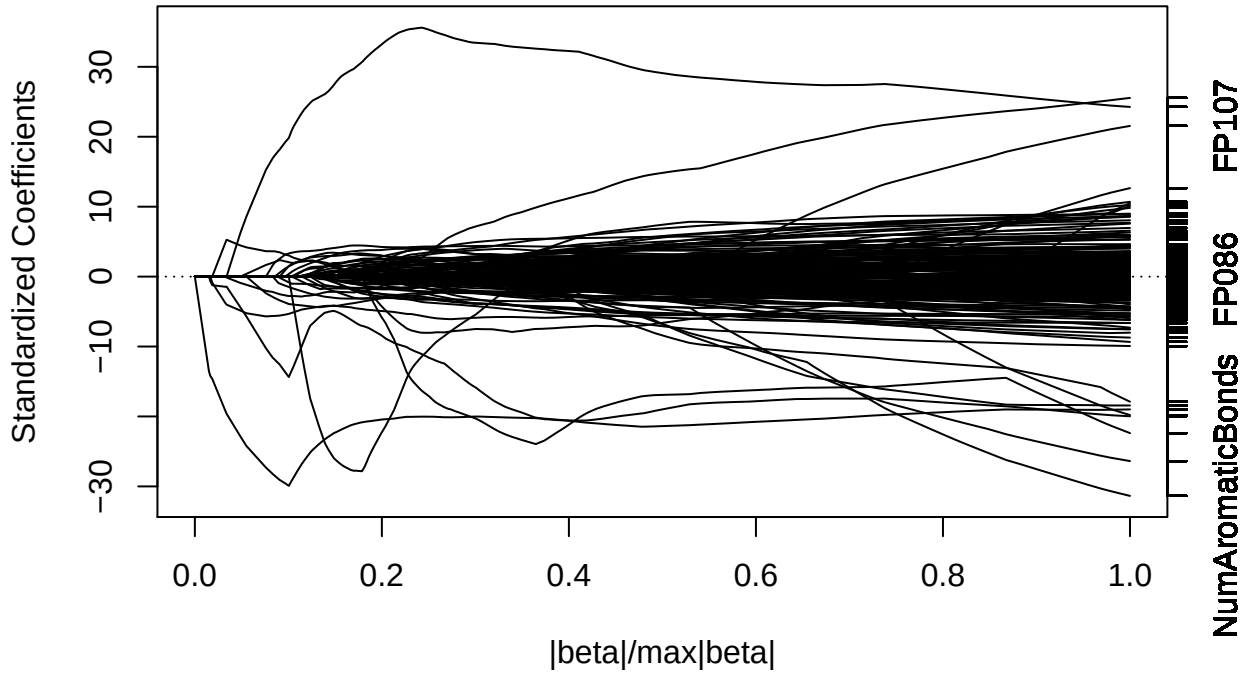
Penalized regression models

```
library(elasticnet)
```

```
## Loading required package: lars
```

```
## Loaded lars 1.2
```

```
ridgeModel <- enet(x = as.matrix(solTrainXtrans), y = solTrainY, lambda = 0.001)
plot(ridgeModel)
```




```
ridgePred <- predict(ridgeModel, newx = as.matrix(solTestXtrans), s=1, mode = "fraction", type = "fit")
head(ridgePred$fit)
```

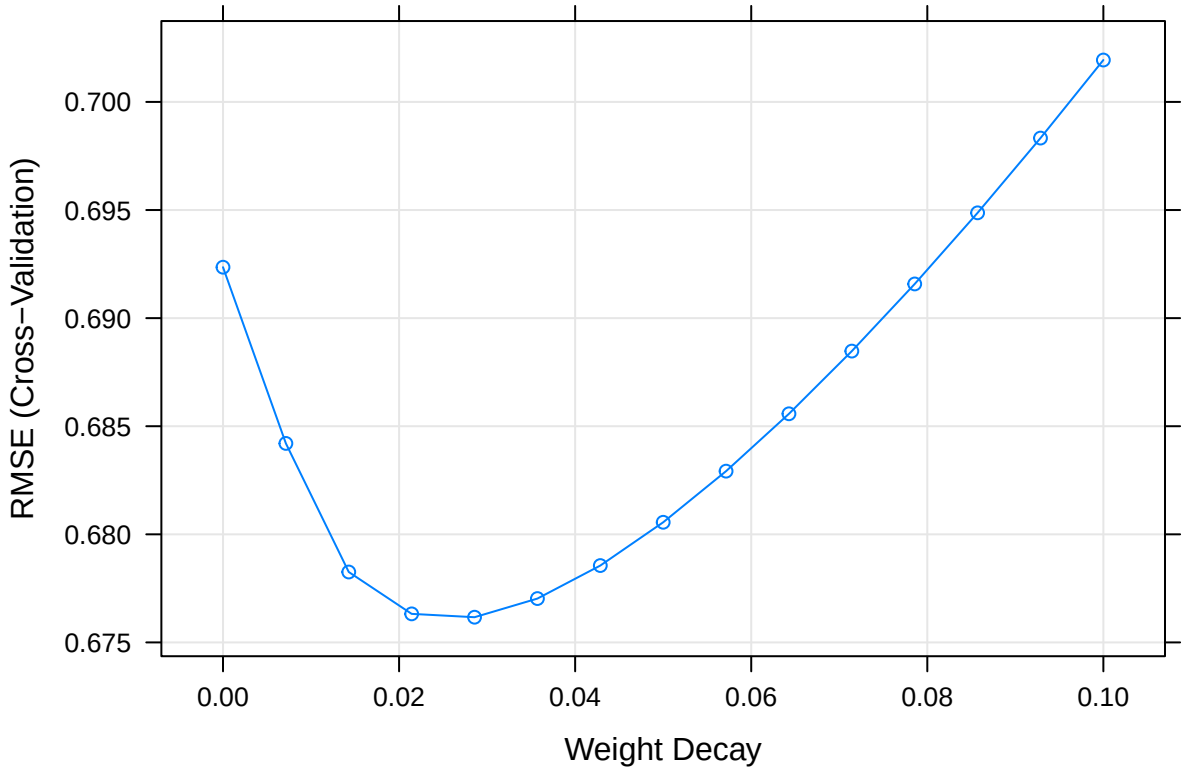
```
##          20          21          23          25          28          31
##  0.96795590  0.06918538 -0.54365077  0.96072014 -0.03594693  1.59284535
```

```
ridgeGrid <- data.frame(.lambda = seq(0, .1, length = 15))
set.seed(100)
ridgeRegFit <- train(solTrainXtrans, solTrainY, method = "ridge", tuneGrid = ridgeGrid,
                     # Fit the model over many penalty values
                     trControl = ctrl, preProc = c("center","scale"))
ridgeRegFit
```

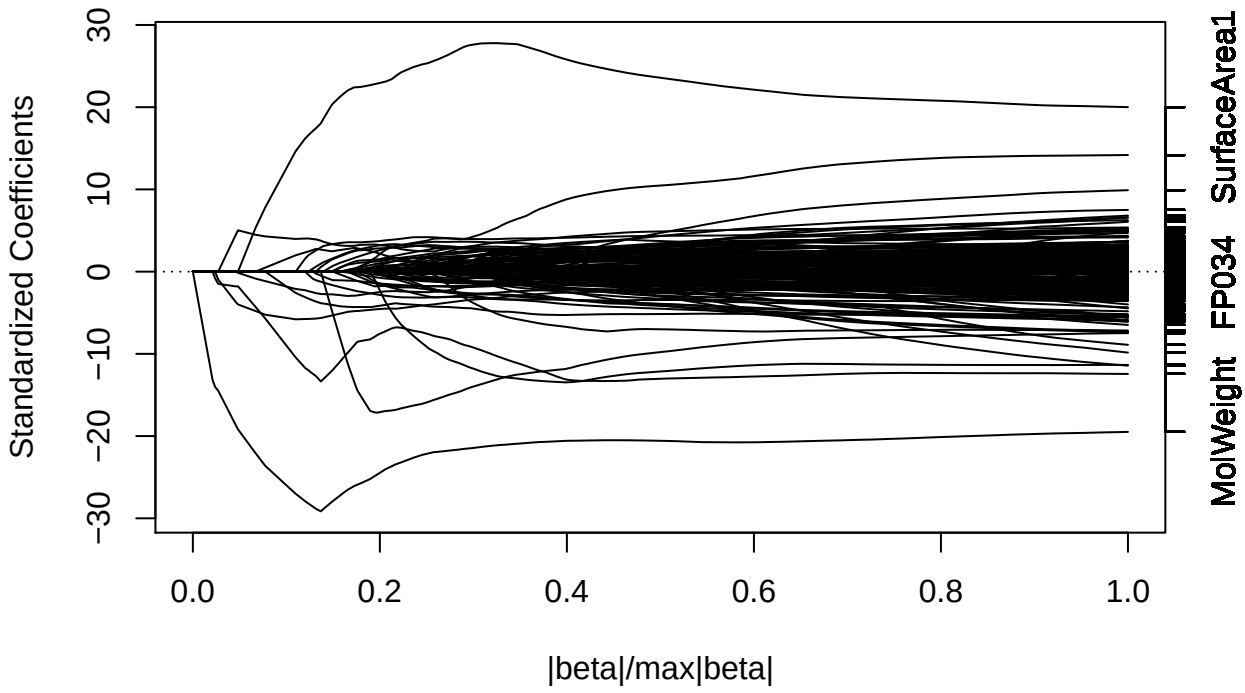
```
## Ridge Regression
##
## 951 samples
## 228 predictors
```

```
##
## Pre-processing: centered (228), scaled (228)
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 856, 856, 855, 855, 857, 856, ...
## Resampling results across tuning parameters:
##
##   lambda       RMSE       Rsquared   MAE
##   0.000000000  0.6923558  0.8872977  0.5194817
##   0.007142857  0.6842051  0.8901855  0.5180204
##   0.014285714  0.6782572  0.8924345  0.5135023
##   0.021428571  0.6763196  0.8933364  0.5129646
##   0.028571429  0.6761659  0.8936611  0.5137609
##   0.035714286  0.6770285  0.8936769  0.5150076
##   0.042857143  0.6785555  0.8935075  0.5169778
##   0.050000000  0.6805575  0.8932196  0.5190373
##   0.057142857  0.6829220  0.8928530  0.5213703
##   0.064285714  0.6855755  0.8924331  0.5238093
##   0.071428571  0.6884742  0.8919761  0.5263585
##   0.078571429  0.6915802  0.8914943  0.5290529
##   0.085714286  0.6948706  0.8909958  0.5318508
##   0.092857143  0.6983276  0.8904864  0.5347012
##   0.100000000  0.7019378  0.8899703  0.5375828
##
## RMSE was used to select the optimal model using the smallest value.
## The final value used for the model was lambda = 0.02857143.
```

```
plot(ridgeRegFit)
```



```
enetModel <- enet(x = as.matrix(solTrainXtrans), y = solTrainY, lambda = 0.01, normalize = TRUE)
plot(enetModel)
```




```
enetPred <- predict(enetModel, newx = as.matrix(solTestXtrans), s = .1, mode = "fraction", type = "fit")
names(enetPred)
```

```
## [1] "s"        "fraction" "mode"     "fit"
```

```
head(enetPred$fit)
```

```
##          20          21          23          25          28          31
## -0.60186178 -0.42226814 -1.20465564 -1.23652963 -1.25023517 -0.05587631
```

```
enetCoef <- predict(enetModel, newx = as.matrix(solTestXTrans), s = .1, mode = "fraction", type = "coef")
tail(enetCoef$coefficients)
```

```
##       NumChlorine        NumHalogen          NumRings HydrophilicFactor
##        0.00000000        0.00000000        0.00000000        0.12678967
##      SurfaceArea1      SurfaceArea2
##        0.09035596        0.00000000
```

```
enetGrid <- expand.grid(.lambda = c(0, 0.01, 0.1), .fraction = seq(0.05, 1, length = 20))
set.seed(100)
enetTune <- train(solTrainXtrans, solTrainY, method = "enet", tuneGrid = enetGrid, trControl = ctrl, pr
```

```
summary(enetGrid)
```

```
##     .lambda           .fraction
##  Min.   :0.00000   Min.   :0.0500
##  1st Qu.:0.00000   1st Qu.:0.2875
##  Median :0.01000   Median :0.5250
##  Mean   :0.03667   Mean   :0.5250
##  3rd Qu.:0.10000   3rd Qu.:0.7625
##  Max.   :0.10000   Max.   :1.0000
```

```
plot(enetTune)
```