HW7

Yigao Li

April 1, 2018

6.8 - 11

(a)

```
library(MASS)
library(glmnet)
```

```
## Warning: package 'glmnet' was built under R version 3.4.3
```

```
## Loading required package: Matrix
```

```
## Loading required package: foreach
```

```
## Warning: package 'foreach' was built under R version 3.4.3
```

```
## Loaded glmnet 2.0-13
```

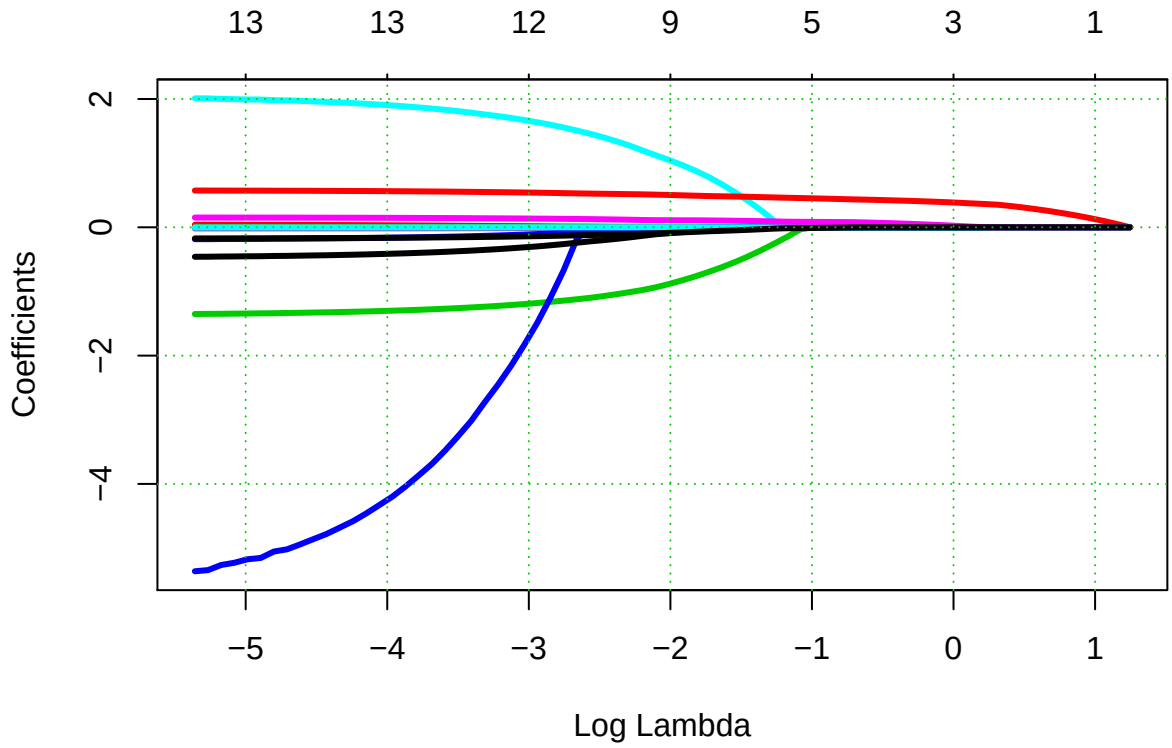
```
boston <- Boston
Split <- floor(dim(boston[1])*0.78)
train <- 1:Split
```

```
## Warning in 1:Split: numerical expression has 2 elements: only the first
## used
```

```
test <- (Split+1):dim(boston)[1]
```

```
## Warning in (Split + 1):dim(boston)[1]: numerical expression has 2 elements:
## only the first used
```

```
boston.train <- boston[train,]
boston.test <- boston[test,]
xtrain.matrix <- model.matrix(crim ~ ., data = boston.train)
xtest.matrix <- model.matrix(crim ~ ., data = boston.test)
y <- boston.train$crim
lasso.fit <- glmnet(xtrain.matrix, y, alpha = 1)
plot(lasso.fit, xvar = "lambda", lwd = 3)
grid(col = 3)
```



(b)

```
lasso.cv <- cv.glmnet(xtrain.matrix, y, alpha = 1)
best.lambda.lasso <- lasso.cv$lambda.min
pred.lasso <- predict(lasso.fit, s = best.lambda.lasso, newx = xtest.matrix)
mean((boston.test$crim - pred.lasso)^2)
```

```
## [1] 128.4983
```

```
coef(lasso.cv, s = best.lambda.lasso)
```

```
## 15 x 1 sparse Matrix of class "dgCMatrix"
##                          1
## (Intercept) -6.922786605
## (Intercept)  .
## zn           0.002797302
## indus        .
## chas        -0.826770860
## nox          .
## rm           0.970130402
## age          .
## dis         -0.074724415
## rad          0.498749895
## tax          0.001691876
## ptratio     -0.019402715
```

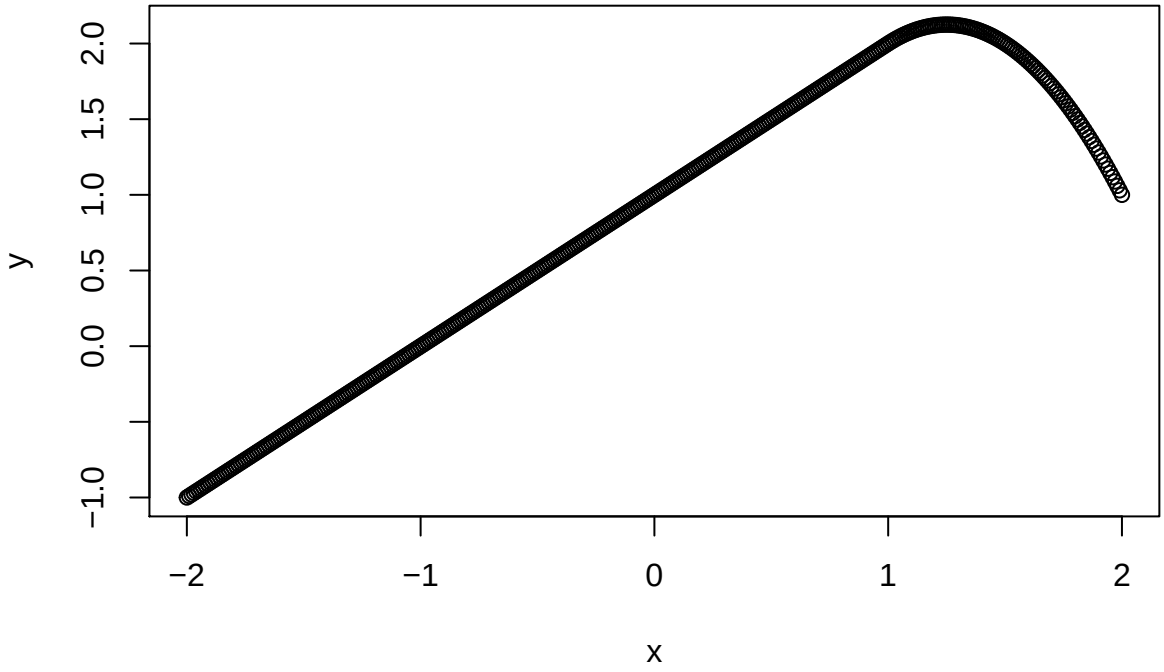
```
## black         .
## lstat         0.109598343
## medv         -0.078744769
```

(c)

No, not all variables are used in LASSO because it forces some coefficients to be set to zero.

7.9 - 3

```
x <- seq(-2,2,0.01)
y <- 1 + x - 2 * (x-1)^2 * I(x>=1)
plot(x,y)
```



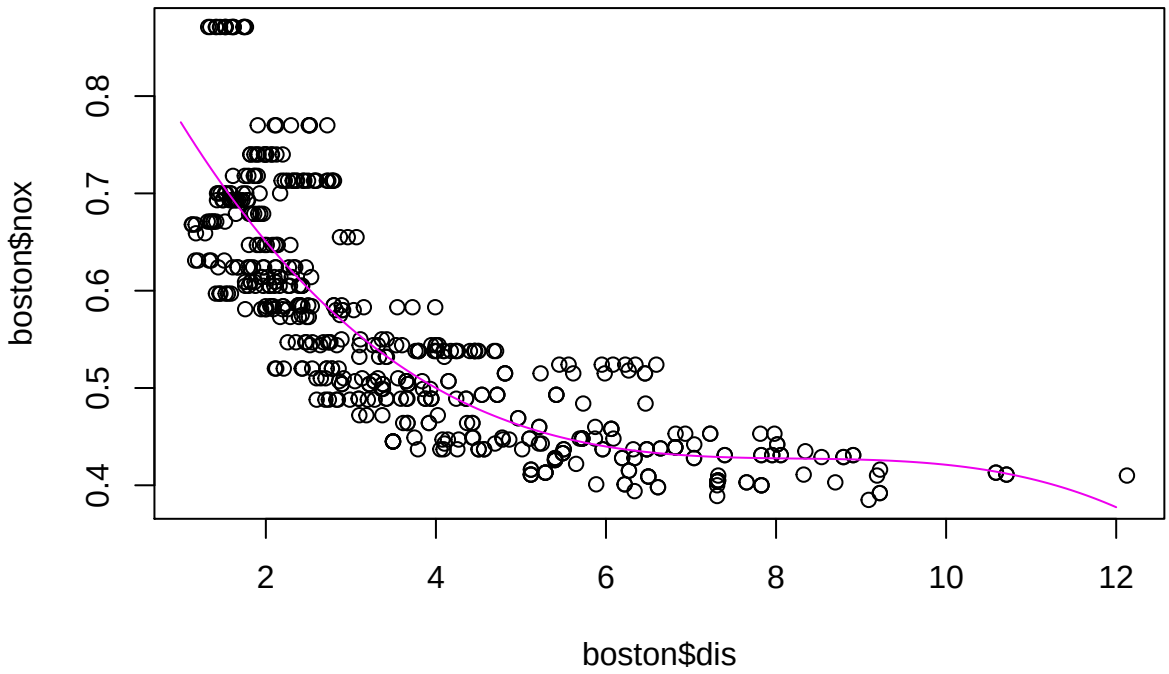
7.9 - 9

(a)

```
model <- lm(nox ~ poly(dis, 3), data = boston)
summary(model)
```

```
##
## Call:
## lm(formula = nox ~ poly(dis, 3), data = boston)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.121130 -0.040619 -0.009738  0.023385  0.194904
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.554695   0.002759 201.021  < 2e-16 ***
## poly(dis, 3)1 -2.003096   0.062071 -32.271  < 2e-16 ***
## poly(dis, 3)2  0.856330   0.062071  13.796  < 2e-16 ***
## poly(dis, 3)3 -0.318049   0.062071  -5.124 4.27e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.06207 on 502 degrees of freedom
## Multiple R-squared:  0.7148, Adjusted R-squared:  0.7131
## F-statistic: 419.3 on 3 and 502 DF,  p-value: < 2.2e-16
```

```r
plot(boston$dis, boston$nox)
dis <- seq(1,12,0.01)
xdf <- data.frame(dis, dis^2, dis^3)
y <- predict(model, xdf)
lines(dis, y, col = "magenta2")
```



(b)

```
for (i in 1:10){
  polyfit <- lm(nox ~ poly(dis, i), data = boston)
  print(sum(polyfit$residual^2))
}
```

```
## [1] 2.768563
## [1] 2.035262
## [1] 1.934107
## [1] 1.932981
## [1] 1.91529
## [1] 1.878257
## [1] 1.849484
## [1] 1.83563
## [1] 1.833331
## [1] 1.832171
```

(c)

```
set.seed(6)
library(boot)
for (i in 1:10){
  polygfit <- glm(nox ~ poly(dis, i), data = boston)
  loocv <- cv.glm(boston, polygfit)
  print(loocv$delta)
}
```

```
## [1] 0.005523868 0.005523815
## [1] 0.004079449 0.004079390
## [1] 0.003874762 0.003874708
## [1] 0.003887521 0.003887449
## [1] 0.004164865 0.004164231
## [1] 0.005384278 0.005381096
## [1] 0.01106878 0.01105424
## [1] 0.008121397 0.008112597
## [1] 0.01761636 0.01758879
## [1] 0.004430276 0.004428769
```

The minimum prediction error is when polynomial degree is 3. Therefore, the optimal degree is 3.

7.9 - 10

(a)

```
library(ISLR)
```

```
## Warning: package 'ISLR' was built under R version 3.4.3
```

```
library(leaps)
```

```
## Warning: package 'leaps' was built under R version 3.4.3
```

```r
college <- College
min.model <- lm(Outstate ~ 1, data = college)
full.model <- lm(Outstate ~ ., data = college)
step(min.model, scope = list(lower = min.model, upper = full.model), direction = "forward")
```

```
## Start:  AIC=12898.87
## Outstate ~ 1
##
##               Df  Sum of Sq        RSS   AIC
## + Expend       1 5684728234 6.8746e+09 12433
## + Room.Board   1 5376025290 7.1833e+09 12467
## + Grad.Rate    1 4099005307 8.4603e+09 12594
## + perc.alumni  1 4027178047 8.5321e+09 12600
## + Top10perc    1 3971446292 8.5879e+09 12606
## + S.F.Ratio    1 3866086440 8.6932e+09 12615
## + Private      1 3835884599 8.7234e+09 12618
## + Top25perc    1 3008031144 9.5513e+09 12688
## + Terminal     1 2090498711 1.0469e+10 12759
## + PhD          1 1842141521 1.0717e+10 12778
## + Personal     1 1123466496 1.1436e+10 12828
## + P.Undergrad  1  807167148 1.1752e+10 12849
## + F.Undergrad  1  584567601 1.1975e+10 12864
## + Enroll       1  303598468 1.2256e+10 12882
## <none>                      1.2559e+10 12899
## + Apps         1   31598296 1.2528e+10 12899
## + Books        1   18960781 1.2540e+10 12900
## + Accept       1    8330540 1.2551e+10 12900
##
## Step:  AIC=12432.62
## Outstate ~ Expend
##
##               Df  Sum of Sq        RSS   AIC
## + Private      1 1930753277 4943815915 12178
## + Room.Board   1 1683455774 5191113418 12216
## + Grad.Rate    1 1411765652 5462803540 12256
## + perc.alumni  1 1237783587 5636785605 12280
## + Personal     1  689772774 6184796418 12352
## + F.Undergrad  1  654774545 6219794647 12357
## + S.F.Ratio    1  500249678 6374319514 12376
## + Enroll       1  497657120 6376912072 12376
## + P.Undergrad  1  492284188 6382285004 12377
## + Top25perc    1  314950268 6559618924 12398
## + Top10perc    1  308836638 6565732554 12399
## + Apps         1  208701510 6665867682 12411
## + Terminal     1  197798147 6676771045 12412
## + Accept       1  153419952 6721149240 12417
## + PhD          1  130314044 6744255148 12420
## <none>                      6874569192 12433
## + Books        1   17199346 6857369846 12433
##
## Step:  AIC=12178.45
## Outstate ~ Expend + Private
##
##               Df Sum of Sq        RSS   AIC
```

```
## + Room.Board   1 957521619 3986294296 12013
## + Terminal     1 754304546 4189511369 12052
## + Grad.Rate    1 724055966 4219759949 12057
## + PhD          1 688814006 4255001909 12064
## + perc.alumni  1 447619509 4496196406 12107
## + Top25perc    1 397739454 4546076461 12115
## + Top10perc    1 323275816 4620540099 12128
## + Personal     1 199132763 4744683152 12148
## + Accept       1 164450738 4779365177 12154
## + Apps         1 115156477 4828659438 12162
## + Enroll       1  29686275 4914129640 12176
## + S.F.Ratio    1  22893877 4920922038 12177
## <none>                     4943815915 12178
## + F.Undergrad  1  11818940 4931996975 12179
## + P.Undergrad  1   8049077 4935766838 12179
## + Books        1   3883230 4939932685 12180
##
## Step:  AIC=12013.18
## Outstate ~ Expend + Private + Room.Board
##
##               Df Sum of Sq        RSS   AIC
## + perc.alumni  1 459760795 3526533501 11920
## + Grad.Rate    1 406258441 3580035855 11932
## + PhD          1 375099014 3611195282 11938
## + Terminal     1 373461127 3612833169 11939
## + Top25perc    1 279816168 3706478128 11959
## + Top10perc    1 255226620 3731067676 11964
## + Personal     1 116422238 3869872058 11992
## + Accept       1  44658844 3941635452 12006
## + P.Undergrad  1  37213229 3949081067 12008
## + Books        1  25622787 3960671509 12010
## + S.F.Ratio    1  24301498 3961992798 12010
## + Apps         1  17189184 3969105112 12012
## <none>                     3986294296 12013
## + Enroll       1   7189256 3979105040 12014
## + F.Undergrad  1    204861 3986089435 12015
##
## Step:  AIC=11919.97
## Outstate ~ Expend + Private + Room.Board + perc.alumni
##
##               Df Sum of Sq        RSS   AIC
## + PhD          1 223766309 3302767192 11871
## + Terminal     1 215610907 3310922594 11873
## + Grad.Rate    1 185543461 3340990040 11880
## + Top25perc    1 118673928 3407859573 11895
## + Top10perc    1 105706952 3420826549 11898
## + Accept       1  64645426 3461888076 11908
## + Personal     1  46442640 3480090861 11912
## + Apps         1  27049039 3499484463 11916
## + Books        1  10686621 3515846880 11920
## + P.Undergrad  1  10149425 3516384076 11920
## + Enroll       1   9960934 3516572567 11920
## + S.F.Ratio    1   9765339 3516768162 11920
## <none>                     3526533501 11920
```

```
## + F.Undergrad  1   2355584 3524177917 11921
##
## Step:  AIC=11871.03
## Outstate ~ Expend + Private + Room.Board + perc.alumni + PhD
##
##               Df Sum of Sq        RSS   AIC
## + Grad.Rate    1 129273253 3173493939 11842
## + Personal     1  48144801 3254622391 11862
## + Top25perc    1  33522465 3269244727 11865
## + Top10perc    1  31577086 3271190106 11866
## + Accept       1  27303953 3275463239 11867
## + Terminal     1  24459320 3278307873 11867
## + P.Undergrad  1  20843690 3281923502 11868
## + S.F.Ratio    1  17212529 3285554663 11869
## <none>                     3302767192 11871
## + Apps         1   7796385 3294970807 11871
## + Books        1   6336668 3296430524 11872
## + F.Undergrad  1   2209904 3300557288 11872
## + Enroll       1     71469 3302695723 11873
##
## Step:  AIC=11842.01
## Outstate ~ Expend + Private + Room.Board + perc.alumni + PhD +
##     Grad.Rate
##
##               Df Sum of Sq        RSS   AIC
## + Personal     1  32007866 3141486073 11836
## + Terminal     1  27622869 3145871070 11837
## + S.F.Ratio    1  19460763 3154033176 11839
## + Accept       1  11686406 3161807533 11841
## + Top25perc    1   9214061 3164279878 11842
## + Top10perc    1   8384772 3165109167 11842
## <none>                     3173493939 11842
## + P.Undergrad  1   7931779 3165562160 11842
## + F.Undergrad  1   5652800 3167841139 11843
## + Books        1   4940188 3168553751 11843
## + Enroll       1   1058307 3172435632 11844
## + Apps         1    218732 3173275207 11844
##
## Step:  AIC=11836.13
## Outstate ~ Expend + Private + Room.Board + perc.alumni + PhD +
##     Grad.Rate + Personal
##
##               Df Sum of Sq        RSS   AIC
## + Terminal     1  26960561 3114525512 11831
## + S.F.Ratio    1  21488428 3119997645 11833
## + Accept       1  16663556 3124822517 11834
## + Top25perc    1  10904040 3130582033 11835
## + Top10perc    1  10080019 3131406054 11836
## <none>                     3141486073 11836
## + P.Undergrad  1   3056303 3138429770 11837
## + F.Undergrad  1   1578173 3139907900 11838
## + Books        1   1415073 3140071000 11838
## + Apps         1   1285269 3140200804 11838
## + Enroll       1      1228 3141484845 11838
```

```
##
## Step:  AIC=11831.43
## Outstate ~ Expend + Private + Room.Board + perc.alumni + PhD +
##     Grad.Rate + Personal + Terminal
##
##               Df Sum of Sq        RSS   AIC
## + S.F.Ratio    1  20590180 3093935332 11828
## + Accept       1  14554658 3099970855 11830
## + Top10perc    1  10079178 3104446334 11831
## + Top25perc    1   8304260 3106221252 11831
## <none>                     3114525512 11831
## + P.Undergrad  1   3795882 3110729631 11832
## + Books        1   3419067 3111106445 11833
## + F.Undergrad  1   2665838 3111859674 11833
## + Apps         1    965869 3113559643 11833
## + Enroll       1     99749 3114425763 11833
##
## Step:  AIC=11828.28
## Outstate ~ Expend + Private + Room.Board + perc.alumni + PhD +
##     Grad.Rate + Personal + Terminal + S.F.Ratio
##
##               Df Sum of Sq        RSS   AIC
## + Accept       1  18802632 3075132700 11826
## + Top10perc    1   9962656 3083972676 11828
## + Top25perc    1   8224868 3085710464 11828
## <none>                     3093935332 11828
## + P.Undergrad  1   2969569 3090965762 11830
## + Books        1   2926489 3091008843 11830
## + Apps         1   2330380 3091604952 11830
## + F.Undergrad  1   1083561 3092851770 11830
## + Enroll       1     55569 3093879763 11830
##
## Step:  AIC=11825.54
## Outstate ~ Expend + Private + Room.Board + perc.alumni + PhD +
##     Grad.Rate + Personal + Terminal + S.F.Ratio + Accept
##
##               Df Sum of Sq        RSS   AIC
## + F.Undergrad  1  64549767 3010582933 11811
## + Enroll       1  52734104 3022398596 11814
## + Apps         1  41326268 3033806432 11817
## + P.Undergrad  1   9100368 3066032332 11825
## <none>                     3075132700 11826
## + Top10perc    1   7725972 3067406728 11826
## + Top25perc    1   5217065 3069915635 11826
## + Books        1   3778233 3071354467 11827
##
## Step:  AIC=11811.06
## Outstate ~ Expend + Private + Room.Board + perc.alumni + PhD +
##     Grad.Rate + Personal + Terminal + S.F.Ratio + Accept + F.Undergrad
##
##               Df Sum of Sq        RSS   AIC
## + Apps         1  43194200 2967388733 11802
## + Top10perc    1  17164307 2993418626 11809
## + Top25perc    1  11926675 2998656258 11810
```

```
## <none>                     3010582933 11811
## + Books         1   2778487 3007804447 11812
## + Enroll        1   1816254 3008766679 11813
## + P.Undergrad   1    329602 3010253331 11813
##
## Step:  AIC=11801.83
## Outstate ~ Expend + Private + Room.Board + perc.alumni + PhD +
##     Grad.Rate + Personal + Terminal + S.F.Ratio + Accept + F.Undergrad +
##     Apps
##
##               Df Sum of Sq        RSS   AIC
## + Top10perc    1  44900435 2922488298 11792
## + Top25perc    1  24513579 2942875154 11797
## <none>                     2967388733 11802
## + Enroll       1   4741299 2962647434 11803
## + Books        1   1990742 2965397991 11803
## + P.Undergrad  1    288766 2967099967 11804
##
## Step:  AIC=11791.98
## Outstate ~ Expend + Private + Room.Board + perc.alumni + PhD +
##     Grad.Rate + Personal + Terminal + S.F.Ratio + Accept + F.Undergrad +
##     Apps + Top10perc
##
##               Df Sum of Sq        RSS   AIC
## + Enroll       1   8945433 2913542865 11792
## <none>                     2922488298 11792
## + Books        1   4132382 2918355916 11793
## + Top25perc    1    361870 2922126428 11794
## + P.Undergrad  1    276700 2922211599 11794
##
## Step:  AIC=11791.6
## Outstate ~ Expend + Private + Room.Board + perc.alumni + PhD +
##     Grad.Rate + Personal + Terminal + S.F.Ratio + Accept + F.Undergrad +
##     Apps + Top10perc + Enroll
##
##               Df Sum of Sq        RSS   AIC
## <none>                     2913542865 11792
## + Books        1   4112173 2909430692 11792
## + Top25perc    1    845842 2912697023 11793
## + P.Undergrad  1    125055 2913417810 11794
##
## Call:
## lm(formula = Outstate ~ Expend + Private + Room.Board + perc.alumni +
##     PhD + Grad.Rate + Personal + Terminal + S.F.Ratio + Accept +
##     F.Undergrad + Apps + Top10perc + Enroll, data = college)
##
## Coefficients:
## (Intercept)       Expend   PrivateYes   Room.Board  perc.alumni
##  -1817.0401       0.2004    2256.9455       0.8742      41.9501
##         PhD    Grad.Rate     Personal     Terminal    S.F.Ratio
##     12.6951      23.8320      -0.2478      22.9730     -47.0056
##      Accept  F.Undergrad         Apps    Top10perc       Enroll
##      0.8024      -0.0957      -0.2999      23.6553      -0.5373
```

(b)

```
library(gam)
```

```
## Warning: package 'gam' was built under R version 3.4.4
```

```
## Loading required package: splines
```

```
## Loaded gam 1.15
```

```
GAM <- gam(Outstate ~ s(Expend, df = 1) + Private + s(Room.Board, df = 5) + s(perc.alumni, df = 3)
           + s(PhD, df = 1) + s(Grad.Rate, df = 5) + s(Personal, df = 2) + s(Terminal, df = 5)
           + s(S.F.Ratio, df = 4) + s(Accept, df = 1) + s(F.Undergrad, df = 1) + s(Apps, df = 1)
           + s(Top10perc, df = 5) + s(Enroll, df = 1), data = college)
summary(GAM)
```

```
##
## Call: gam(formula = Outstate ~ s(Expend, df = 1) + Private + s(Room.Board,
##     df = 5) + s(perc.alumni, df = 3) + s(PhD, df = 1) + s(Grad.Rate,
##     df = 5) + s(Personal, df = 2) + s(Terminal, df = 5) + s(S.F.Ratio,
##     df = 4) + s(Accept, df = 1) + s(F.Undergrad, df = 1) + s(Apps,
##     df = 1) + s(Top10perc, df = 5) + s(Enroll, df = 1), data = college)
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -6695.43 -1225.92   -32.41  1278.83  9503.91
##
## (Dispersion Parameter for gaussian family taken to be 3547626)
##
##     Null Deviance: 12559297426 on 776 degrees of freedom
## Residual Deviance: 2625227130 on 739.9955 degrees of freedom
## AIC: 13961.67
##
## Number of Local Scoring Iterations: 2
##
## Anova for Parametric Effects
##                           Df     Sum Sq    Mean Sq   F value    Pr(>F)
## s(Expend, df = 1)          1 5805174796 5805174796 1636.3549 < 2.2e-16 ***
## Private                    1 1728385723 1728385723  487.1951 < 2.2e-16 ***
## s(Room.Board, df = 5)      1  921085142  921085142  259.6342 < 2.2e-16 ***
## s(perc.alumni, df = 3)     1  387400810  387400810  109.2000 < 2.2e-16 ***
## s(PhD, df = 1)             1  197859484  197859484   55.7724 2.295e-13 ***
## s(Grad.Rate, df = 5)       1  142721584  142721584   40.2302 3.928e-10 ***
## s(Personal, df = 2)        1   29664081   29664081    8.3617  0.003945 **
## s(Terminal, df = 5)        1   27975896   27975896    7.8858  0.005114 **
## s(S.F.Ratio, df = 4)       1   20261176   20261176    5.7112  0.017106 *
## s(Accept, df = 1)          1   17462856   17462856    4.9224  0.026813 *
## s(F.Undergrad, df = 1)     1   64690780   64690780   18.2350 2.207e-05 ***
## s(Apps, df = 1)            1   29809487   29809487    8.4027  0.003858 **
## s(Top10perc, df = 5)       1   38635285   38635285   10.8905  0.001013 **
## s(Enroll, df = 1)          1    4234811    4234811    1.1937  0.274939
## Residuals                740 2625227130    3547626
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Anova for Nonparametric Effects
```

```
##                         Npar Df Npar F     Pr(F)
## (Intercept)
## s(Expend, df = 1)            0 48.176 0.0005853 ***
## Private
## s(Room.Board, df = 5)        4  2.395 0.0490939 *
## s(perc.alumni, df = 3)       2  0.686 0.5037140
## s(PhD, df = 1)               0  3.447 2.448e-05 ***
## s(Grad.Rate, df = 5)         4  2.400 0.0487172 *
## s(Personal, df = 2)          1  5.743 0.0167989 *
## s(Terminal, df = 5)          4  3.061 0.0162070 *
## s(S.F.Ratio, df = 4)         3 10.435 9.902e-07 ***
## s(Accept, df = 1)            0 20.367 0.0012529 **
## s(F.Undergrad, df = 1)       0  4.933 0.0031433 **
## s(Apps, df = 1)              0 16.246 0.0029753 **
## s(Top10perc, df = 5)         4  1.426 0.2234982
## s(Enroll, df = 1)            0  7.091 0.0006335 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```