Hw6p4

Sagar Ganapaneni

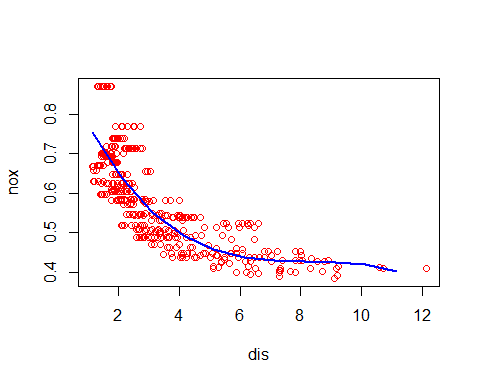
November 15, 2016

1. Use the "poly()" function to fit a cubic polynomial regression to predict "nox" using "dis". Report the regression output, and plot the resulting data and polynomial fits.

set.seed(12)  
par(mfrow = c(1,1))  
  
library(MASS)  
library(boot)  
library(splines)  
df<-Boston  
fit <- lm(nox ~ poly(dis, 3), data = df)  
summary(fit)

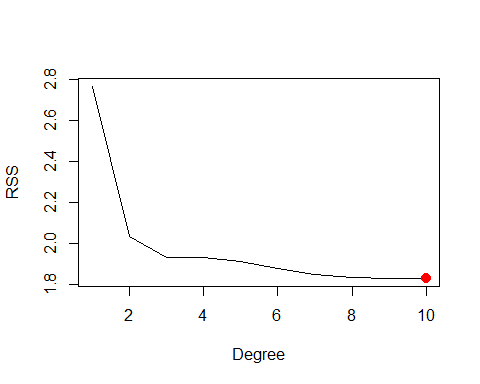
##   
## Call:  
## lm(formula = nox ~ poly(dis, 3), data = df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.121130 -0.040619 -0.009738 0.023385 0.194904   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.554695 0.002759 201.021 < 2e-16 \*\*\*  
## poly(dis, 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

dis\_range = range(df$dis)  
dis\_seq = seq(from = dis\_range[1], to = dis\_range[2])  
prediction = predict(fit, list(dis = dis\_seq))  
plot(nox ~ dis, data = df,col="red")  
lines(dis\_seq, prediction, lwd = 2, col = 'blue')

  
We may conclude that all polynomial terms are significant.

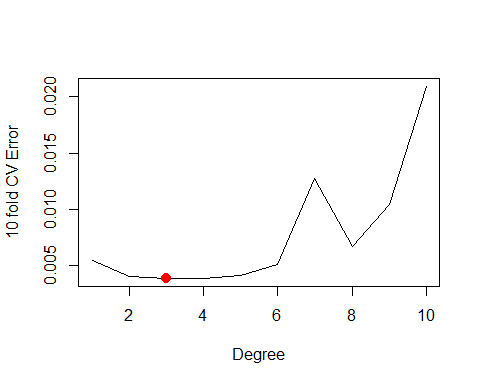
1. Plot the polynomial fits for a range of different polynomial degrees (say, from 1 to 10), and report the associated residual sum of squares.

RSS <- rep(NA, 10)  
for (i in 1:10) {  
 fit <- lm(nox ~ poly(dis, i), data = Boston)  
 RSS[i] <- sum(fit$residuals^2)  
}  
plot(1:10, RSS, xlab = "Degree", ylab = "RSS", type = "l")  
points(which.min(RSS), RSS[which.min(RSS)], col = "red", cex = 2, pch = 20)

  
It seems that the RSS decreases with the degree of the polynomial, and so is minimum for a polynomial of degree 10.

1. Perform cross-validation or another approach to select the optimal degree for the polynomial, and explain your results.

MSE <- rep(NA, 10)  
for (i in 1:10) {  
 fit <- glm(nox ~ poly(dis, i), data = Boston)  
 MSE[i] <- cv.glm(Boston, fit, K = 10)$delta[1]  
}  
plot(1:10, MSE, xlab = "Degree", ylab = "10 fold CV Error", type = "l")  
points(which.min(MSE), MSE[which.min(MSE)], col = "red", cex = 2, pch = 20)

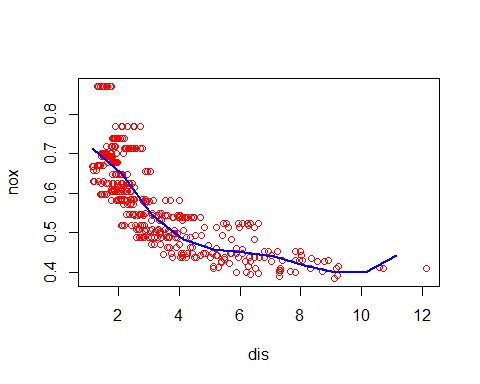
  
We may see that a polynomial of degree 4 minimizes the test MSE.

1. Use the "bs()" function to fit a regression spline to predict "nox" using "dis". Report the output for the fit using 7 degrees of freedom (3 knots)

fit <- lm(nox ~ bs(dis, df = 4, knots = c(3, 7, 11)), data = df)  
summary(fit)

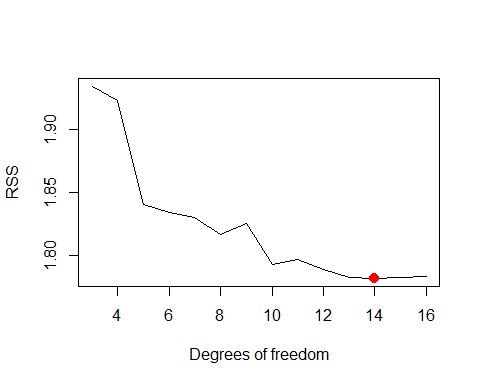
##   
## Call:  
## lm(formula = nox ~ bs(dis, df = 4, knots = c(3, 7, 11)), data = df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.130710 -0.039850 -0.008357 0.027792 0.188518   
##   
## Coefficients:  
## Estimate Std. Error t value  
## (Intercept) 0.714346 0.015846 45.081  
## bs(dis, df = 4, knots = c(3, 7, 11))1 -0.006626 0.024307 -0.273  
## bs(dis, df = 4, knots = c(3, 7, 11))2 -0.296980 0.018293 -16.234  
## bs(dis, df = 4, knots = c(3, 7, 11))3 -0.222840 0.033763 -6.600  
## bs(dis, df = 4, knots = c(3, 7, 11))4 -0.379811 0.042317 -8.975  
## bs(dis, df = 4, knots = c(3, 7, 11))5 -0.222959 0.086870 -2.567  
## bs(dis, df = 4, knots = c(3, 7, 11))6 -0.304346 0.063378 -4.802  
## Pr(>|t|)   
## (Intercept) < 2e-16 \*\*\*  
## bs(dis, df = 4, knots = c(3, 7, 11))1 0.7853   
## bs(dis, df = 4, knots = c(3, 7, 11))2 < 2e-16 \*\*\*  
## bs(dis, df = 4, knots = c(3, 7, 11))3 1.05e-10 \*\*\*  
## bs(dis, df = 4, knots = c(3, 7, 11))4 < 2e-16 \*\*\*  
## bs(dis, df = 4, knots = c(3, 7, 11))5 0.0106 \*   
## bs(dis, df = 4, knots = c(3, 7, 11))6 2.08e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.06137 on 499 degrees of freedom  
## Multiple R-squared: 0.7229, Adjusted R-squared: 0.7196   
## F-statistic: 217 on 6 and 499 DF, p-value: < 2.2e-16

pred <- predict(fit, list(dis = dis\_seq))  
plot(nox ~ dis, data = Boston, col = "red")  
lines(dis\_seq, pred, col = "blue", lwd = 2)

  
I chose the knots based such that the data was roughly split into even pieces (since dis 's lower limit is ~1 and its upper limit is ~13).

1. Now fit a regression spline for a range of degrees of freedom, and plot the resulting fits and report the resulting RSS. Describe the results obtained.

RSS <- rep(NA, 16)  
for (i in 3:16) {  
 fit <- lm(nox ~ bs(dis, df = i), data = Boston)  
 RSS[i] <- sum(fit$residuals^2)  
}  
plot(3:16, RSS[-c(1, 2)], xlab = "Degrees of freedom", ylab = "RSS", type = "l")  
points(which.min(RSS), RSS[which.min(RSS)], col = "red", cex = 2, pch = 20)

  
We may see that RSS decreases until 14 and then slightly increases after that.

1. Perform cross-validation or another approach in order to select the best degrees of freedom for a regression spline on this data. Describe your results.

CVError <- rep(NA, 16)  
for (i in 3:16) {  
 fit <- glm(nox ~ bs(dis, df = i), data = Boston)  
 CVError[i] <- cv.glm(Boston, fit, K = 10)$delta[1]  
}

## Warning in bs(dis, degree = 3L, knots = numeric(0), Boundary.knots =  
## c(1.1296, : some 'x' values beyond boundary knots may cause ill-conditioned  
## bases  
  
## Warning in bs(dis, degree = 3L, knots = numeric(0), Boundary.knots =  
## c(1.1296, : some 'x' values beyond boundary knots may cause ill-conditioned  
## bases

## Warning in bs(dis, degree = 3L, knots = numeric(0), Boundary.knots =  
## c(1.137, : some 'x' values beyond boundary knots may cause ill-conditioned  
## bases  
  
## Warning in bs(dis, degree = 3L, knots = numeric(0), Boundary.knots =  
## c(1.137, : some 'x' values beyond boundary knots may cause ill-conditioned  
## bases

## Warning in bs(dis, degree = 3L, knots = structure(3.1675, .Names =  
## "50%"), : some 'x' values beyond boundary knots may cause ill-conditioned  
## bases  
  
## Warning in bs(dis, degree = 3L, knots = structure(3.1675, .Names =  
## "50%"), : some 'x' values beyond boundary knots may cause ill-conditioned  
## bases

## Warning in bs(dis, degree = 3L, knots = structure(3.1827, .Names =  
## "50%"), : some 'x' values beyond boundary knots may cause ill-conditioned  
## bases  
  
## Warning in bs(dis, degree = 3L, knots = structure(3.1827, .Names =  
## "50%"), : some 'x' values beyond boundary knots may cause ill-conditioned  
## bases

## Warning in bs(dis, degree = 3L, knots = structure(c(2.3727,  
## 4.36263333333333: some 'x' values beyond boundary knots may cause ill-  
## conditioned bases  
  
## Warning in bs(dis, degree = 3L, knots = structure(c(2.3727,  
## 4.36263333333333: some 'x' values beyond boundary knots may cause ill-  
## conditioned bases

## Warning in bs(dis, degree = 3L, knots = structure(c(2.4212,  
## 4.23913333333333: some 'x' values beyond boundary knots may cause ill-  
## conditioned bases  
  
## Warning in bs(dis, degree = 3L, knots = structure(c(2.4212,  
## 4.23913333333333: some 'x' values beyond boundary knots may cause ill-  
## conditioned bases

## Warning in bs(dis, degree = 3L, knots = structure(c(2.10035, 3.2157,  
## 5.16495: some 'x' values beyond boundary knots may cause ill-conditioned  
## bases  
  
## Warning in bs(dis, degree = 3L, knots = structure(c(2.10035, 3.2157,  
## 5.16495: some 'x' values beyond boundary knots may cause ill-conditioned  
## bases

## Warning in bs(dis, degree = 3L, knots = structure(c(2.0754, 3.1323,  
## 5.11735: some 'x' values beyond boundary knots may cause ill-conditioned  
## bases  
  
## Warning in bs(dis, degree = 3L, knots = structure(c(2.0754, 3.1323,  
## 5.11735: some 'x' values beyond boundary knots may cause ill-conditioned  
## bases

## Warning in bs(dis, degree = 3L, knots = structure(c(1.96376, 2.66502,  
## 3.9175, : some 'x' values beyond boundary knots may cause ill-conditioned  
## bases  
  
## Warning in bs(dis, degree = 3L, knots = structure(c(1.96376, 2.66502,  
## 3.9175, : some 'x' values beyond boundary knots may cause ill-conditioned  
## bases

## Warning in bs(dis, degree = 3L, knots = structure(c(1.9512, 2.6403,  
## 3.9454, : some 'x' values beyond boundary knots may cause ill-conditioned  
## bases  
  
## Warning in bs(dis, degree = 3L, knots = structure(c(1.9512, 2.6403,  
## 3.9454, : some 'x' values beyond boundary knots may cause ill-conditioned  
## bases

## Warning in bs(dis, degree = 3L, knots = structure(c(1.8651,  
## 2.41306666666667, : some 'x' values beyond boundary knots may cause ill-  
## conditioned bases  
  
## Warning in bs(dis, degree = 3L, knots = structure(c(1.8651,  
## 2.41306666666667, : some 'x' values beyond boundary knots may cause ill-  
## conditioned bases

## Warning in bs(dis, degree = 3L, knots = structure(c(1.86156666666667,  
## 2.38403333333333, : some 'x' values beyond boundary knots may cause ill-  
## conditioned bases  
  
## Warning in bs(dis, degree = 3L, knots = structure(c(1.86156666666667,  
## 2.38403333333333, : some 'x' values beyond boundary knots may cause ill-  
## conditioned bases

## Warning in bs(dis, degree = 3L, knots = structure(c(1.78037142857143,  
## 2.2044, : some 'x' values beyond boundary knots may cause ill-conditioned  
## bases  
  
## Warning in bs(dis, degree = 3L, knots = structure(c(1.78037142857143,  
## 2.2044, : some 'x' values beyond boundary knots may cause ill-conditioned  
## bases

## Warning in bs(dis, degree = 3L, knots = structure(c(1.81317142857143,  
## 2.25881428571429, : some 'x' values beyond boundary knots may cause ill-  
## conditioned bases  
  
## Warning in bs(dis, degree = 3L, knots = structure(c(1.81317142857143,  
## 2.25881428571429, : some 'x' values beyond boundary knots may cause ill-  
## conditioned bases

## Warning in bs(dis, degree = 3L, knots = structure(c(1.743225, 2.0754,  
## 2.4999, : some 'x' values beyond boundary knots may cause ill-conditioned  
## bases  
  
## Warning in bs(dis, degree = 3L, knots = structure(c(1.743225, 2.0754,  
## 2.4999, : some 'x' values beyond boundary knots may cause ill-conditioned  
## bases

## Warning in bs(dis, degree = 3L, knots = structure(c(1.6732,  
## 2.00496666666667, : some 'x' values beyond boundary knots may cause ill-  
## conditioned bases  
  
## Warning in bs(dis, degree = 3L, knots = structure(c(1.6732,  
## 2.00496666666667, : some 'x' values beyond boundary knots may cause ill-  
## conditioned bases

## Warning in bs(dis, degree = 3L, knots = structure(c(1.6723,  
## 2.00613333333333, : some 'x' values beyond boundary knots may cause ill-  
## conditioned bases  
  
## Warning in bs(dis, degree = 3L, knots = structure(c(1.6723,  
## 2.00613333333333, : some 'x' values beyond boundary knots may cause ill-  
## conditioned bases

## Warning in bs(dis, degree = 3L, knots = structure(c(1.6362, 1.9865,  
## 2.288, : some 'x' values beyond boundary knots may cause ill-conditioned  
## bases  
  
## Warning in bs(dis, degree = 3L, knots = structure(c(1.6362, 1.9865,  
## 2.288, : some 'x' values beyond boundary knots may cause ill-conditioned  
## bases

## Warning in bs(dis, degree = 3L, knots = structure(c(1.64668, 1.96376,  
## 2.28422, : some 'x' values beyond boundary knots may cause ill-conditioned  
## bases  
  
## Warning in bs(dis, degree = 3L, knots = structure(c(1.64668, 1.96376,  
## 2.28422, : some 'x' values beyond boundary knots may cause ill-conditioned  
## bases

## Warning in bs(dis, degree = 3L, knots = structure(c(1.60816363636364,  
## 1.87607272727273, : some 'x' values beyond boundary knots may cause ill-  
## conditioned bases  
  
## Warning in bs(dis, degree = 3L, knots = structure(c(1.60816363636364,  
## 1.87607272727273, : some 'x' values beyond boundary knots may cause ill-  
## conditioned bases

## Warning in bs(dis, degree = 3L, knots = structure(c(1.59590909090909,  
## 1.87607272727273, : some 'x' values beyond boundary knots may cause ill-  
## conditioned bases  
  
## Warning in bs(dis, degree = 3L, knots = structure(c(1.59590909090909,  
## 1.87607272727273, : some 'x' values beyond boundary knots may cause ill-  
## conditioned bases

## Warning in bs(dis, degree = 3L, knots = structure(c(1.58948333333333,  
## 1.8301, : some 'x' values beyond boundary knots may cause ill-conditioned  
## bases  
  
## Warning in bs(dis, degree = 3L, knots = structure(c(1.58948333333333,  
## 1.8301, : some 'x' values beyond boundary knots may cause ill-conditioned  
## bases

## Warning in bs(dis, degree = 3L, knots = structure(c(1.60973333333333,  
## 1.8651, : some 'x' values beyond boundary knots may cause ill-conditioned  
## bases  
  
## Warning in bs(dis, degree = 3L, knots = structure(c(1.60973333333333,  
## 1.8651, : some 'x' values beyond boundary knots may cause ill-conditioned  
## bases

## Warning in bs(dis, degree = 3L, knots = structure(c(1.5523,  
## 1.79772307692308, : some 'x' values beyond boundary knots may cause ill-  
## conditioned bases  
  
## Warning in bs(dis, degree = 3L, knots = structure(c(1.5523,  
## 1.79772307692308, : some 'x' values beyond boundary knots may cause ill-  
## conditioned bases

## Warning in bs(dis, degree = 3L, knots = structure(c(1.5895, 1.8195,  
## 2.0407, : some 'x' values beyond boundary knots may cause ill-conditioned  
## bases  
  
## Warning in bs(dis, degree = 3L, knots = structure(c(1.5895, 1.8195,  
## 2.0407, : some 'x' values beyond boundary knots may cause ill-conditioned  
## bases

## Warning in bs(dis, degree = 3L, knots = structure(c(1.5311,  
## 1.78037142857143, : some 'x' values beyond boundary knots may cause ill-  
## conditioned bases  
  
## Warning in bs(dis, degree = 3L, knots = structure(c(1.5311,  
## 1.78037142857143, : some 'x' values beyond boundary knots may cause ill-  
## conditioned bases

plot(3:16, CVError[-c(1, 2)], xlab = "Degrees of freedom", ylab = "10 fold CV Error", type = "l")

