R Lab 7

Jared Andreatta

2025-04-24

7.8.1 Polynomial Regression and Step Functions

```
library(ISLR2)

# Fitting with orthogonal polynomials of degree 4
fit <- lm(wage~poly(age,4), data=Wage)
fit.summary <- summary(fit)
coef(fit.summary)
```

```
##                 Estimate Std. Error    t value     Pr(>|t|)
## (Intercept)    111.70361  0.7287409 153.283015 0.000000e+00
## poly(age, 4)1  447.06785 39.9147851  11.200558 1.484604e-28
## poly(age, 4)2 -478.31581 39.9147851 -11.983424 2.355831e-32
## poly(age, 4)3  125.52169 39.9147851   3.144742 1.678622e-03
## poly(age, 4)4  -77.91118 39.9147851  -1.951938 5.103865e-02
```

```
# We can get polynomials directly with raw=TRUE
fit.raw <- lm(wage~poly(age, 4, raw=TRUE), data=Wage)
fit.raw.summary <- summary(fit.raw)
coef(fit.raw.summary)
```

```
##                                Estimate   Std. Error   t value     Pr(>|t|)
## (Intercept)               -1.841542e+02 6.004038e+01 -3.067172 0.0021802539
## poly(age, 4, raw = TRUE)1  2.124552e+01 5.886748e+00  3.609042 0.0003123618
## poly(age, 4, raw = TRUE)2 -5.638593e-01 2.061083e-01 -2.735743 0.0062606446
## poly(age, 4, raw = TRUE)3  6.810688e-03 3.065931e-03  2.221409 0.0263977518
## poly(age, 4, raw = TRUE)4 -3.203830e-05 1.641359e-05 -1.951938 0.0510386498
```

```
# Fitting with I() (same as doing raw polynomial)
fit.I <- lm(wage ~ age + I(age^2) + I(age^3) + I(age^4), data = Wage)
fit.I.summary <- summary(fit.I)
coef(fit.I.summary)
```

```
##                  Estimate   Std. Error   t value     Pr(>|t|)
## (Intercept) -1.841542e+02 6.004038e+01 -3.067172 0.0021802539
## age          2.124552e+01 5.886748e+00  3.609042 0.0003123618
## I(age^2)    -5.638593e-01 2.061083e-01 -2.735743 0.0062606446
## I(age^3)     6.810688e-03 3.065931e-03  2.221409 0.0263977518
## I(age^4)    -3.203830e-05 1.641359e-05 -1.951938 0.0510386498
```

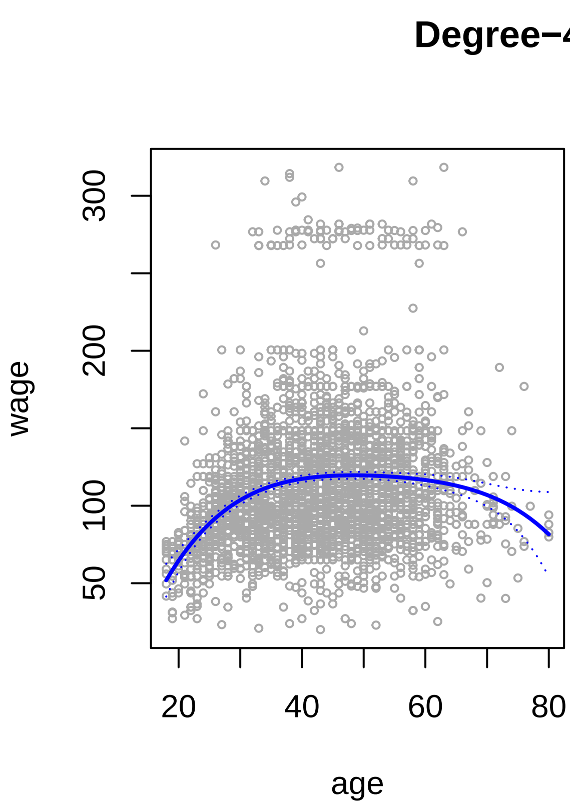
```
# Doing same thing with cbind()
fit.cbind <- lm(wage ~ cbind(age, age^2, age^3, age^4), data = Wage)

# Creating grid of values to make predictions
```

```
attach(Wage)
# Age grid
agelims <- range(age)
age.grid <- seq(from = agelims[1], to = agelims[2])

# Predictions
preds <- predict(fit, newdata = list(age = age.grid),
se = TRUE)

# Standard error bands
se.bands <- cbind(preds$fit + 2 * preds$se.fit,
preds$fit - 2 * preds$se.fit)

# Plotting
par(mfrow = c(1, 2), mar = c(4.5, 4.5, 1, 1), oma = c(0, 0, 4, 0))
plot(age, wage, xlim = agelims, cex = .5, col = "darkgrey") > title("Degree-4 Polynomial", outer = T)
```

```
## logical(0)
```

```
lines(age.grid, preds$fit, lwd = 2, col = "blue")
matlines(age.grid, se.bands, lwd = 1, col = "blue", lty = 3)

#
preds2 <- predict(fit.raw, newdata = list(age = age.grid), se = TRUE)
max(abs(preds$fit - preds2$fit))
```

```
## [1] 6.88658e-11
```



ANOVA

```r
# Testing different degrees of polynomials to assess model performance
# Results indicate either cubic or quartic polynomials
fit.1 <- lm(wage ~ age, data = Wage)
fit.2 <- lm(wage ~ poly(age, 2), data = Wage)
fit.3 <- lm(wage ~ poly(age, 3), data = Wage)
fit.4 <- lm(wage ~ poly(age, 4), data = Wage)
fit.5 <- lm(wage ~ poly(age, 5), data = Wage)
anova(fit.1, fit.2, fit.3, fit.4, fit.5)
```

```
## Analysis of Variance Table
##
## Model 1: wage ~ age
## Model 2: wage ~ poly(age, 2)
## Model 3: wage ~ poly(age, 3)
## Model 4: wage ~ poly(age, 4)
## Model 5: wage ~ poly(age, 5)
##   Res.Df     RSS Df Sum of Sq        F    Pr(>F)
## 1   2998 5022216
## 2   2997 4793430  1    228786 143.5931 < 2.2e-16 ***
## 3   2996 4777674  1     15756   9.8888  0.001679 **
## 4   2995 4771604  1      6070   3.8098  0.051046 .
## 5   2994 4770322  1      1283   0.8050  0.369682
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Just using the 5th degree polynomial model to check performance
coef(summary(fit.5))
```

```
##                 Estimate Std. Error     t value     Pr(>|t|)
## (Intercept)    111.70361  0.7287647 153.2780243 0.000000e+00
## poly(age, 5)1  447.06785 39.9160847  11.2001930 1.491111e-28
## poly(age, 5)2 -478.31581 39.9160847 -11.9830341 2.367734e-32
## poly(age, 5)3  125.52169 39.9160847   3.1446392 1.679213e-03
## poly(age, 5)4  -77.91118 39.9160847  -1.9518743 5.104623e-02
## poly(age, 5)5  -35.81289 39.9160847  -0.8972045 3.696820e-01
```

```r
# Works whether or not we do orthogonal polynomials
fit.1 <- lm(wage ~ education + age, data = Wage)
fit.2 <- lm(wage ~ education + poly(age, 2), data = Wage)
fit.3 <- lm(wage ~ education + poly(age, 3), data = Wage)
anova(fit.1, fit.2, fit.3)
```

```
## Analysis of Variance Table
##
## Model 1: wage ~ education + age
## Model 2: wage ~ education + poly(age, 2)
## Model 3: wage ~ education + poly(age, 3)
##   Res.Df     RSS Df Sum of Sq        F Pr(>F)
## 1   2994 3867992
## 2   2993 3725395  1    142597 114.6969 <2e-16 ***
## 3   2992 3719809  1      5587   4.4936 0.0341 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```