Stat 482, Homework #9 Solutions

1. y is vibration, x_1 is bit size, x_2 is cutting speed

(a) An interaction effect occurs when the effect of an input depends on the levels of the other inputs.

(b) $b_1 = 8.3187$, $b_2 = 3.7687$, $b_{12} = 4.3562$, $SE = 0.6112$

(c) $x_2 = +1, \quad \hat{y} = 27.6 + 12.675x_1$

$x_2 = 0, \quad \hat{y} = 23.8 + 8.319x_1$

$x_2 = -1, \quad \hat{y} = 20.1 + 3.9625x_1$

(d) see attached for interaction plots

(e) $t^*_{12} = 7.127$, $p = .000$ The observed data is not compatible with the additive effects model. We accept the model which includes an interaction between size and speed for predicting vibration.

2. (a) quadratic model, $R^2 = .8143$

(b) cubic model, $R^2 = .8273$

(c) see attached for scatterplot

Overfitting a model is more serious when we extrapolate outside the observed input space

HW 9, Problem 1 Computing

Data from Handout

A router is used to cut locating notches on a circuit board. The vibration level is considered to be an important characteristic of the process. Two factors are thought to affect vibration (y): bit size (x1) and cutting speed (x2).

:

```
setwd("F:/Lexar/stat 482 data sets")
hw.dat = read.csv('hw7-1.csv')

hw.dat$x1 = 2*(hw.dat$bit.size - mean(hw.dat$bit.size)) /
(range(hw.dat$bit.size )[2]-range(hw.dat$bit.size )[1])

hw.dat$x2 = 2*(hw.dat$cutting.speed - mean(hw.dat$cutting.speed)) /
(range(hw.dat$cutting.speed)[2]-range(hw.dat$cutting.speed)[1])

hw.dat$y = hw.dat$vibration
hw.dat
```

```
##    bit.size cutting.speed vibration x1 x2    y
## 1    0.0625            40      18.2 -1 -1 18.2
## 2    0.1250            40      27.2  1 -1 27.2
## 3    0.0625            90      15.9 -1  1 15.9
## 4    0.1250            90      41.0  1  1 41.0
## 5    0.0625            40      18.9 -1 -1 18.9
## 6    0.1250            40      24.0  1 -1 24.0
## 7    0.0625            90      14.5 -1  1 14.5
## 8    0.1250            90      43.9  1  1 43.9
## 9    0.0625            40      12.9 -1 -1 12.9
## 10   0.1250            40      22.4  1 -1 22.4
## 11   0.0625            90      15.1 -1  1 15.1
## 12   0.1250            90      36.3  1  1 36.3
## 13   0.0625            40      14.4 -1 -1 14.4
## 14   0.1250            40      22.5  1 -1 22.5
## 15   0.0625            90      14.2 -1  1 14.2
## 16   0.1250            90      39.9  1  1 39.9
```

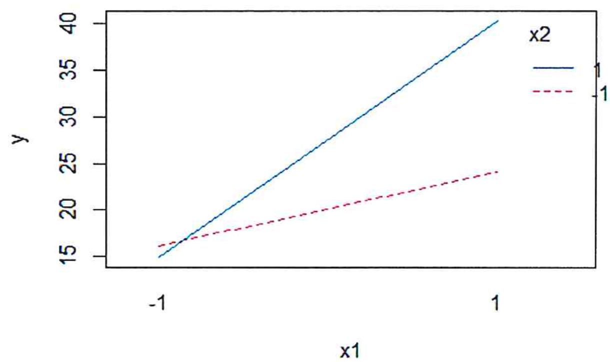
```r
interaction.mod = lm(y ~ x1+x2+I(x1*x2),data=hw.dat)
summary(interaction.mod)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2 + I(x1 * x2), data = hw.dat)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.975 -1.550 -0.200  1.256  3.625
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  23.8312     0.6112  38.991 5.22e-14 ***
## x1            8.3187     0.6112  13.611 1.17e-08 ***
## x2            3.7687     0.6112   6.166 4.83e-05 ***
## I(x1 * x2)    4.3562     0.6112   7.127 1.20e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.445 on 12 degrees of freedom
## Multiple R-squared:  0.9581, Adjusted R-squared:  0.9476
## F-statistic: 91.36 on 3 and 12 DF,  p-value: 1.569e-08
```

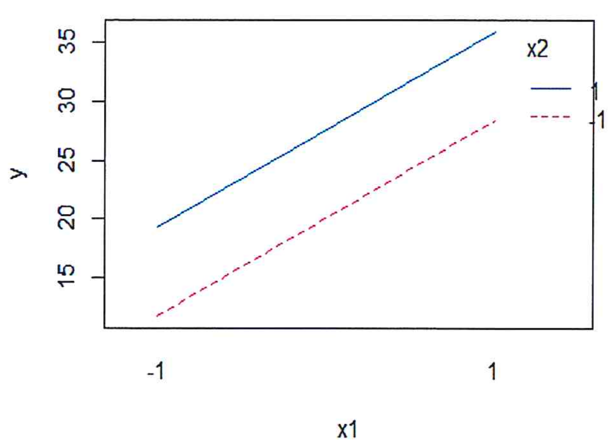
```r
b0 = interaction.mod$coefficients[1]
b1 = interaction.mod$coefficients[2]
b2 = interaction.mod$coefficients[3]
b12 = interaction.mod$coefficients[4]

reg.estimates.x1 = matrix(c(b0+b2,b0,b0-b2,b1+b12,b1,b1-b12),nrow = 3)
dimnames(reg.estimates.x1) = list(c("x2=+1","x2=0","x2=-
1"),c("intercept","slope"))

reg.estimates.x1
```

```
##       intercept    slope
## x2=+1  27.60000 12.67500
## x2=0   23.83125  8.31875
## x2=-1  20.06250  3.96250
```

```r
interaction.plot(hw.dat$x1,hw.dat$x2,hw.dat$y, col =
c("red","blue","green"),trace.label = "x2",
                 xlab = "x1",ylab = "y")
```




```r
additive.mod = lm(y ~ x1+x2,data=hw.dat)
pred.add = predict(additive.mod)

interaction.plot(hw.dat$x1,hw.dat$x2,pred.add, col =
c("red","blue","green"),trace.label = "x2",
                 xlab = "x1",ylab = "y")
```



HW #9, Problem 2 Computing

Data from Exercise 8.6

A sample of healthy females is selected to investigate the relationship between age (x) and the level of a steroid (y)

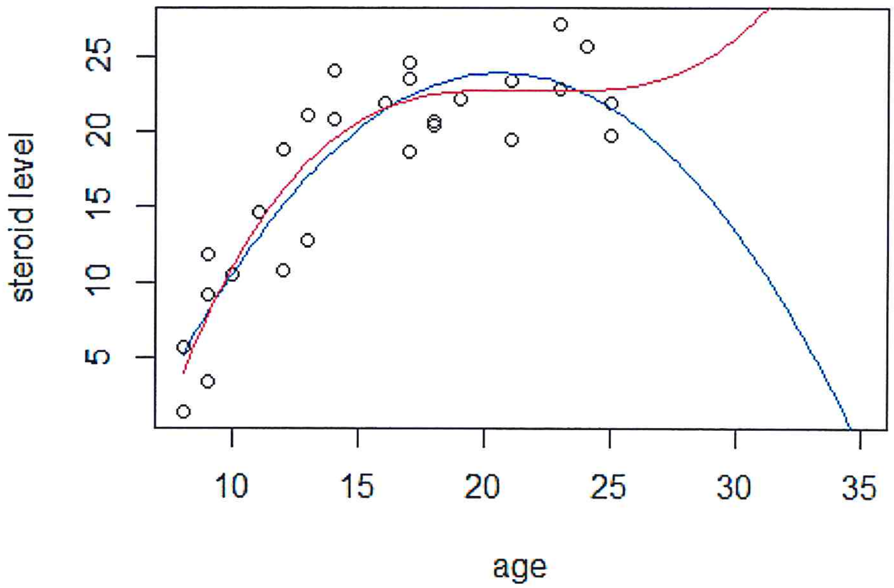
:

```
hw9.data = read.table(
'http://users.stat.ufl.edu/~rrandles/sta4210/Rclassnotes/data/textdatasets/KutnerData/Chapter%20%208%20Data%20Sets/CH08PR06.txt')

colnames(hw9.data)=c("steroid.level","age")
str(hw9.data)

## 'data.frame':    27 obs. of  2 variables:
##  $ steroid.level: num  27.1 22.1 21.9 10.7 1.4 18.8 14.7 5.7 18.6 20.4 ...
##  $ age          : num  23 19 25 12 8 12 11 8 17 18 ...

quad.mod = lm(steroid.level ~ poly(age,2),data = hw9.data)
summary(quad.mod)

##
## Call:
## lm(formula = steroid.level ~ poly(age, 2), data = hw9.data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.5463 -2.5369  0.3868  2.1973  5.3020
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     17.6444     0.6067  29.081  < 2e-16 ***
## poly(age, 2)1  28.1652     3.1526   8.934 4.24e-09 ***
## poly(age, 2)2 -15.9055     3.1526  -5.045 3.71e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.153 on 24 degrees of freedom
## Multiple R-squared:  0.8143, Adjusted R-squared:  0.7989
## F-statistic: 52.63 on 2 and 24 DF,  p-value: 1.678e-09
```

```r
cubic.mod = lm(steroid.level ~ poly(age,3),data = hw9.data)
summary(cubic.mod)
```

```
##
## Call:
## lm(formula = steroid.level ~ poly(age, 3), data = hw9.data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.3854 -2.2414  0.3715  2.1975  4.5595
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      17.6444     0.5977  29.521  < 2e-16 ***
## poly(age, 3)1    28.1652     3.1057   9.069 4.67e-09 ***
## poly(age, 3)2   -15.9055     3.1057  -5.121 3.45e-05 ***
## poly(age, 3)3     4.0861     3.1057   1.316    0.201
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.106 on 23 degrees of freedom
## Multiple R-squared:  0.8273, Adjusted R-squared:  0.8048
## F-statistic: 36.73 on 3 and 23 DF,  p-value: 6.122e-09
```

```r
grid = data.frame(age = seq(min(hw9.data$age), max(hw9.data$age)+10,
length.out = 100))
grid$quad = predict(quad.mod, newdata = grid)
grid$cubic = predict(cubic.mod, newdata = grid)

plot(hw9.data$age,hw9.data$steroid.level, xlim =
c(min(hw9.data$age),max(hw9.data$age)+10),
     xlab = "age",ylab = "steroid level")
points(grid$age,grid$quad,type="l",col="blue")
points(grid$age,grid$cubic,type="l",col="red")
```




```
accepted.mod = lm(steroid.level ~ age + I(age^2),data=hw9.data)
summary(accepted.mod)

##
## Call:
## lm(formula = steroid.level ~ age + I(age^2), data = hw9.data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.5463 -2.5369  0.3868  2.1973  5.3020
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -26.32541    5.88154  -4.476 0.000157 ***
## age           4.87357    0.77515   6.287 1.69e-06 ***
## I(age^2)     -0.11840    0.02347  -5.045 3.71e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.153 on 24 degrees of freedom
## Multiple R-squared:  0.8143, Adjusted R-squared:  0.7989
## F-statistic: 52.63 on 2 and 24 DF,  p-value: 1.678e-09
```