ST5227_Tut_1

Name:Zhu Xu
User ID:E0337988
Student ID:A0131944H

06/02/2019

Question_1:

$$\hat{\beta}_R = \min_{\beta}\{\sum_{i=1}^{n}(Y_i - X_i^T\beta)^2 + \lambda\sum_{k=1}^{p}\beta_k^2\}$$

Let $Q(\beta_1, \cdots \beta_p) = \sum_{i=1}^{n}\{Y_i - (\beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip})\} + \lambda\sum_{k=1}^{p}\beta_k^2$

The LSE satisfy:

$$\frac{\partial Q(\beta_1,\cdots\beta_p)}{\partial\beta_1} = \frac{\partial Q(\beta_1,\cdots\beta_p)}{\partial\beta_2} = \cdots = \frac{\partial Q(\beta_1,\cdots\beta_p)}{\partial\beta_p} = 0$$

In matrix:

$$X^TX\hat{\beta}_R + \lambda I\hat{\beta}_R = X^TY \quad \Rightarrow \hat{\beta}_R = (X^TX + \lambda I)^{-1}X^TY$$

For $\hat{\beta}_R = \min_{\beta}\{\sum_{i=1}^{n}(Y_i - X_i^T\beta)^2 + \sum_{k=1}^{p}\lambda_k\beta_k^2\}$

Let $Q_1(\beta_1, \cdots \beta_p) = \sum_{i=1}^{n}(Y_i - X_i^T\beta)^2 + \sum_{k=1}^{p}\lambda_k\beta_k^2$

$$\frac{\partial Q_1(\beta_1,\cdots\beta_p)}{\partial\beta_1} = \frac{\partial Q_1(\beta_1,\cdots\beta_p)}{\partial\beta_2} = \cdots = \frac{\partial Q_1(\beta_1,\cdots\beta_p)}{\partial\beta_p} = 0$$

In matrix:

$$X^TX\hat{\beta}_R + diag(\lambda_k)\hat{\beta}_R = X^TY \quad \Rightarrow \hat{\beta}_R = [X^TX + diag(\lambda_k)]^{-1}X^TY$$

Question_2:

$$3x_1 - 2x_2 + x_3 = \begin{pmatrix}3 & -2 & 1\end{pmatrix}\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix} = ax$$

$3x_1 - 2x_2 + x_3$ follows normal distribution and:

$$\mu = E(ax) = aE(x) = \begin{pmatrix}3 & -2 & 1\end{pmatrix}\begin{pmatrix}1\\2\\3\end{pmatrix} = 2$$

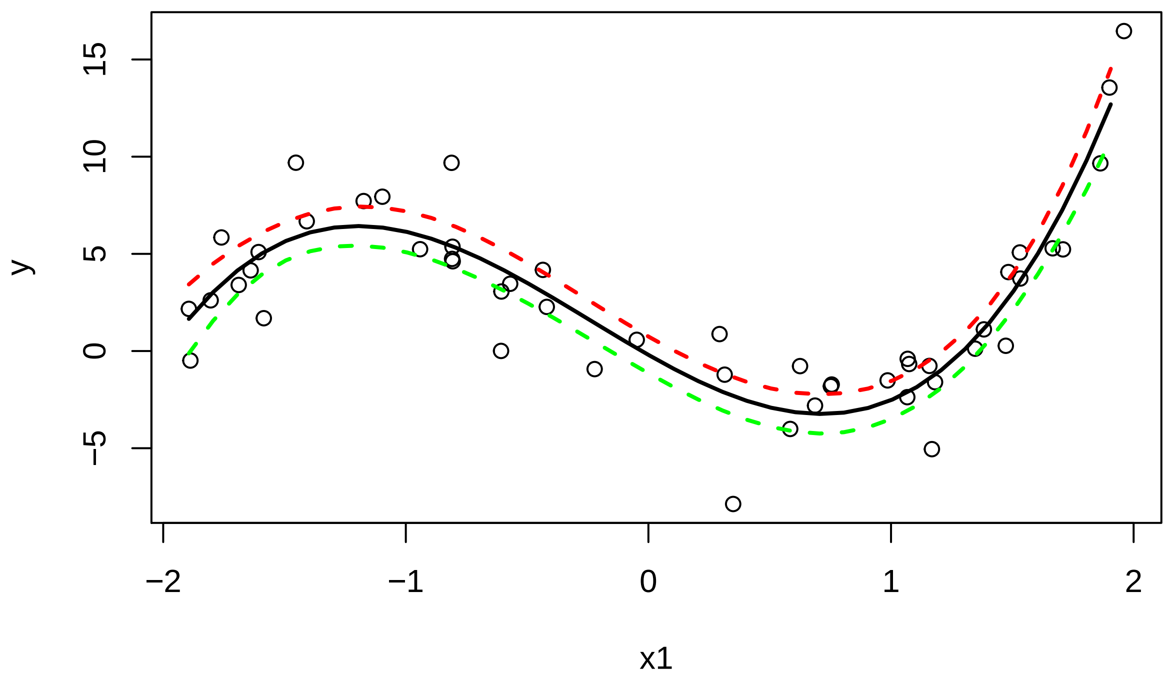
$$Var(ax) = aVar(x)a^T = \begin{pmatrix}3 & -2 & 1\end{pmatrix}\begin{pmatrix}1 & 0.5 & 0.4\\0.5 & 2 & 1\\0.4 & 1 & 3\end{pmatrix}\begin{pmatrix}3\\-2\\1\end{pmatrix} = 12.4$$

Question_3:

```
data =
  read.table(
    "/Users/xuzhu/Desktop/Notes/Sem2/ST5227-Applied_Data_Mining/Tut/Tut1/data01T01.dat"
    )
x1 = data[,1]
y = data[,2]
x2 = x1^2
x3 = x1^3
reg = lm(y~x1+x2+x3)

new_x1 = seq(min(x1), max(x1), 0.1)
new_x2 = new_x1^2
new_x3 = new_x1^3
new_data = data.frame(x1=new_x1, x2=new_x2, x3=new_x3)

pred = predict(reg, newdata=new_data, interval="confidence", level=0.95)
plot(x1,y)
lines(new_x1, pred[,1], lty=1, col="black",lwd=2)
lines(new_x1, pred[,2], lty=2, col="green",lwd=2)
lines(new_x1, pred[,3], lty=2, col="red",lwd=2)
```



Question_4:

Based on BIC:

BIC(145)=-4.6600(the smallest one)

Based on BIC and using forward selection:

BIC(0)=0.9596, BIC(1)=0.7902, BIC(2)=1.0517, BIC(3)=1.0050, BIC(4)=0.7338, BIC(5)=0.7950

Thus we choose x_4 into model.

Then BIC(14)=0.5632, BIC(24)=0.8257, BIC(34)=0.7451, BIC(45)=-0.192

Thus we choose x_5 into model.

Then BIC(145)=-4.6600, BIc(245)=-0.1099, BIC(345)=-0.1223.

Finally we choose model 145.

Based on BIC and using backward selection:

BIC(12345)=-4.5600, BIC(1234)=0.6714, BIC(1235)=0.5204, BIC(1245)=-4.6093, BIC(1345)=-4.6157, BIC(2345)=-0.0372

Thus drop 2 out of the model.

Then BIC(134)=0.5914, BIC(135)=0.4283, BIC(145)=-4.6600

Finally we choose model 145.

Question_5:

$Y_i = \beta X_i + \varepsilon_i$ and $E\hat{Y} \sim N(EY, \frac{\sigma^2 x^2}{\sum_{i=1}^n X_i^2})$

Thus the 95% CI for EY is $[E\hat{Y} - 1.96\sigma\sqrt{\frac{x^2}{\sum_{i=1}^n X_i^2}}, E\hat{Y} + 1.96\sigma\sqrt{\frac{x^2}{\sum_{i=1}^n X_i^2}}]$

Question_6:

$Y - \bar{Y} = a + b_1x_1 + b_2x_2 + \cdots + b_px_p + e - \bar{y} = b_1(x_1 - \bar{x_1}) + b_2(x_2 - \bar{x_2}) + \cdots + b_p(x_p - \bar{x_p}) + e$

$\Rightarrow b_1s_1\frac{x_1-\bar{x}_1}{s1} + b_2s_2\frac{x_2-\bar{x}_2}{s2} + \cdots + b_ps_p\frac{x_p-\bar{x}_p}{sp} + e$

$\Rightarrow b_1s_1x_1' + b_2s_2x_2' + \cdots + b_ps_px_p' + e$

Thus $\frac{Y-\bar{Y}}{s} = Y' = b_1\frac{s_1}{s}x_1' + \cdots + b_p\frac{s_p}{s}x_p' + \frac{e}{s} = b_1'x_1' + \cdots + b_p'x_p' + e'$

$\Rightarrow b_k' = \frac{s_k}{s}b_k$