# 第六次 統計學習作業 學號 M072040015 姓名葉永傑

14. This problem focuses on the *collinearity* problem.

(a) Perform the following commands in R:

```
> set.seed(1)
> x1=runif(100)
> x2=0.5*x1+rnorm(100)/10
> y=2+2*x1+0.3*x2+rnorm(100)
```

The last line corresponds to creating a linear model in which y is a function of x1 and x2. Write out the form of the linear model. What are the regression coefficients?

(b) What is the correlation between x1 and x2? Create a scatterplot displaying the relationship between the variables.

(c) Using this data, fit a least squares regression to predict y using x1 and x2. Describe the results obtained. What are $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\beta}_2$? How do these relate to the true $\beta_0$, $\beta_1$, and $\beta_2$? Can you reject the null hypothesis $H_0 : \beta_1 = 0$? How about the null hypothesis $H_0 : \beta_2 = 0$?

(d) Now fit a least squares regression to predict y using only x1. Comment on your results. Can you reject the null hypothesis $H_0 : \beta_1 = 0$?

(e) Now fit a least squares regression to predict y using only x2. Comment on your results. Can you reject the null hypothesis $H_0 : \beta_1 = 0$?

(f) Do the results obtained in (c)–(e) contradict each other? Explain your answer.

(g) Now suppose we obtain one additional observation, which was unfortunately mismeasured.

```
> x1=c(x1, 0.1)
> x2=c(x2, 0.8)
> y=c(y,6)
```

Re-fit the linear models from (c) to (e) using this new data. What effect does this new observation have on the each of the models? In each model, is this observation an outlier? A high-leverage point? Both? Explain your answers.

## (a)

```
set.seed (1)
x1=runif (100)
x2 =0.5* x1+rnorm (100) /10
y=2+2* x1 +0.3* x2+rnorm (100)
```

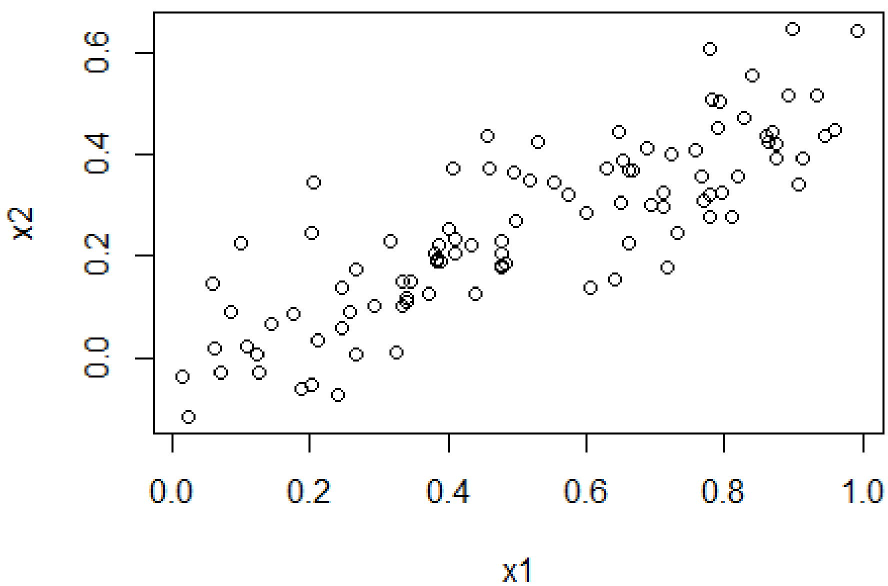
$$y = 2 + 2x_1 + 0.3x_2 + \epsilon\ ,\epsilon\ \mathcal{N}(0,1)$$

## (b)

```
cor(x1,x2)

## [1] 0.8351212

plot(x1,x2)
```




## (c)

```
fit1<-lm(y~x1+x2)
summary(fit1)

##
## Call:
## lm(formula = y ~ x1 + x2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.8311 -0.7273 -0.0537  0.6338  2.3359
```

```
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.1305     0.2319   9.188 7.61e-15 ***
## x1            1.4396     0.7212   1.996   0.0487 *
## x2            1.0097     1.1337   0.891   0.3754
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.056 on 97 degrees of freedom
## Multiple R-squared:  0.2088, Adjusted R-squared:  0.1925
## F-statistic:  12.8 on 2 and 97 DF,  p-value: 1.164e-05
```

$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$

$\hat{y} = 2.1305 + 1.4396x_1 + 1.0097x_2$

reject

$$H_0; \beta_1 = 0$$

,not reject $H_0; \beta_2 = 0$

## (d)

```
fit2<-lm(y~x1)
summary(fit2)
```

```
##
## Call:
## lm(formula = y ~ x1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.89495 -0.66874 -0.07785  0.59221  2.45560
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.1124     0.2307   9.155 8.27e-15 ***
## x1            1.9759     0.3963   4.986 2.66e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.055 on 98 degrees of freedom
## Multiple R-squared:  0.2024, Adjusted R-squared:  0.1942
## F-statistic: 24.86 on 1 and 98 DF,  p-value: 2.661e-06
```

reject

$$H_0; \beta_1 = 0$$

## (e)

```r
fit3<-lm(y~x2)
summary(fit3)
```

```
##
## Call:
## lm(formula = y ~ x2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.62687 -0.75156 -0.03598  0.72383  2.44890
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.3899     0.1949   12.26  < 2e-16 ***
## x2            2.8996     0.6330    4.58 1.37e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.072 on 98 degrees of freedom
## Multiple R-squared:  0.1763, Adjusted R-squared:  0.1679
## F-statistic: 20.98 on 1 and 98 DF,  p-value: 1.366e-05
```

## (f)

reject$H_0; \beta_1 = 0$ 在簡單線性中拒絕的$\beta_1$ 在復回歸中不成立的原因，應該是 x1&x2 共線性的問題，導致 過於相似的向量解釋同一個變數

## (g)

```r
x1 <- c(x1, 0.1)
x2 <- c(x2, 0.8)
y <- c(y, 6)
fit4 <- lm(y ~ x1 + x2)
fit5 <- lm(y ~ x1)
fit6 <- lm(y ~ x2)
summary(fit4)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.73348 -0.69318 -0.05263  0.66385  2.30619
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.2267     0.2314   9.624 7.91e-16 ***
## x1            0.5394     0.5922   0.911  0.36458
```

```
## x2            2.5146     0.8977   2.801  0.00614 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.075 on 98 degrees of freedom
## Multiple R-squared:  0.2188, Adjusted R-squared:  0.2029
## F-statistic: 13.72 on 2 and 98 DF,  p-value: 5.564e-06
```

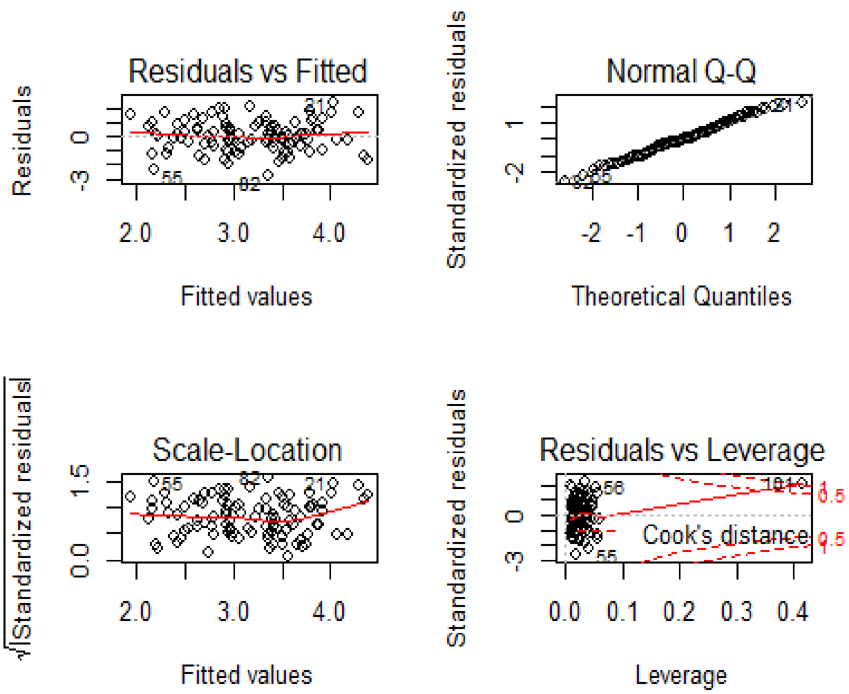
```r
summary(fit5)
```

```
##
## Call:
## lm(formula = y ~ x1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.8897 -0.6556 -0.0909  0.5682  3.5665
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.2569     0.2390   9.445 1.78e-15 ***
## x1            1.7657     0.4124   4.282 4.29e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.111 on 99 degrees of freedom
## Multiple R-squared:  0.1562, Adjusted R-squared:  0.1477
## F-statistic: 18.33 on 1 and 99 DF,  p-value: 4.295e-05
```

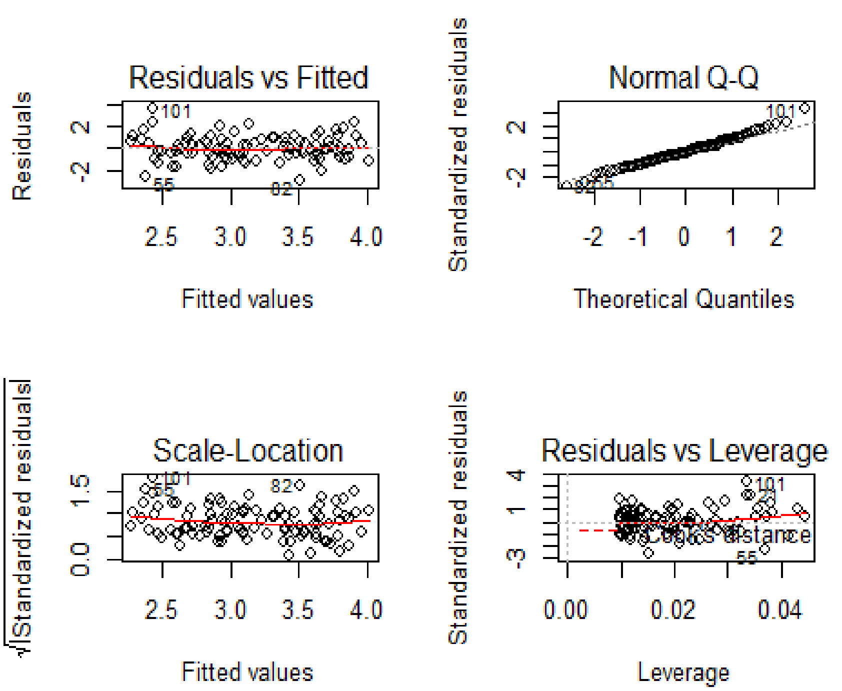
```r
summary(fit6)
```

```
##
## Call:
## lm(formula = y ~ x2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.64729 -0.71021 -0.06899  0.72699  2.38074
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.3451     0.1912  12.264  < 2e-16 ***
## x2            3.1190     0.6040   5.164 1.25e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.074 on 99 degrees of freedom
## Multiple R-squared:  0.2122, Adjusted R-squared:  0.2042
## F-statistic: 26.66 on 1 and 99 DF,  p-value: 1.253e-06
```

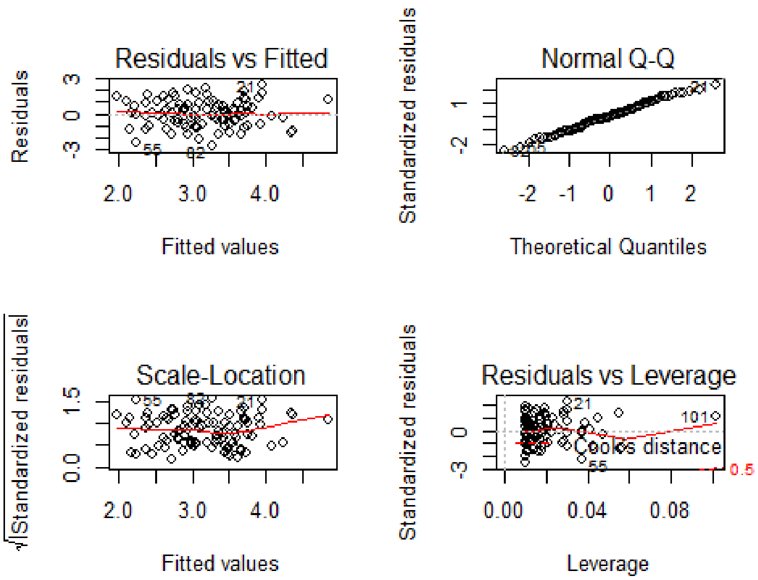
```
par(mfrow=c(2,2))
plot(fit4)
```




```
plot(fit5)
```



```r
plot(fit6)
```




在 y~x1+x2 的模型中，第 101 是高槓桿點。 y~x1 的模型是異常值。 在以 y~x2 的模型中，第 101 是高槓桿點。

15. This problem involves the `Boston` data set, which we saw in the lab for this chapter. We will now try to predict per capita crime rate using the other variables in this data set. In other words, per capita crime rate is the response, and the other variables are the predictors.

(a) For each predictor, fit a simple linear regression model to predict the response. Describe your results. In which of the models is there a statistically significant association between the predictor and the response? Create some plots to back up your assertions.

(b) Fit a multiple regression model to predict the response using all of the predictors. Describe your results. For which predictors can we reject the null hypothesis $H_0 : \beta_j = 0$?

(c) How do your results from (a) compare to your results from (b)? Create a plot displaying the univariate regression coefficients from (a) on the $x$-axis, and the multiple regression coefficients from (b) on the $y$-axis. That is, each predictor is displayed as a single point in the plot. Its coefficient in a simple linear regression model is shown on the $x$-axis, and its coefficient estimate in the multiple linear regression model is shown on the $y$-axis.

(d) Is there evidence of non-linear association between any of the predictors and the response? To answer this question, for each predictor $X$, fit a model of the form

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \epsilon.$$

## (a)

```
data(Boston)
table1<-Map(summary,(Map(function(x)lm(crim~x,data=Boston),Boston[,-
1])))
a=data.frame(13,2)

for(i in 1:13)
{a[i,1]<-Map(function(x)lm(crim~x,data=Boston),Boston[,-1])[[i]]$coef[1]
a[i,2]<-table1[[i]]$coefficients[7]
a[i,3]<-Map(function(x)lm(crim~x,data=Boston),Boston[,-1])[[i]]$coef[2]
a[i,4]<-table1[[i]]$coefficients[8]

}

colnames(a)<-c("b_0","p-value","b_1","p-value")
rownames(a)<-colnames(Boston)[-1]
a

##                b_0      p-value         b_1      p-value
## zn        4.453694 4.037668e-24 -0.07393498 5.506472e-06
```

```
## indus    -2.063743 2.091266e-03  0.50977633 1.450349e-21
## chas      3.744447 1.239505e-19 -1.89277655 2.094345e-01
## nox     -13.719882 5.076814e-15 31.24853120 3.751739e-23
## rm       20.481804 2.272000e-09 -2.68405122 6.346703e-07
## age      -3.777906 7.221718e-05  0.10778623 2.854869e-16
## dis       9.499262 1.502748e-33 -1.55090168 8.519949e-19
## rad      -2.287159 3.605846e-07  0.61791093 2.693844e-56
## tax      -8.528369 2.773600e-23  0.02974225 2.357127e-47
## ptratio -17.646933 3.395255e-08  1.15198279 2.942922e-11
## black    16.553529 8.922239e-28 -0.03627964 2.487274e-19
## lstat    -3.330538 2.087022e-06  0.54880478 2.654277e-27
## medv     11.796536 5.934119e-32 -0.36315992 1.173987e-19
```

只有 chas 的 b_1 沒有 reject $H_0$

## (b)

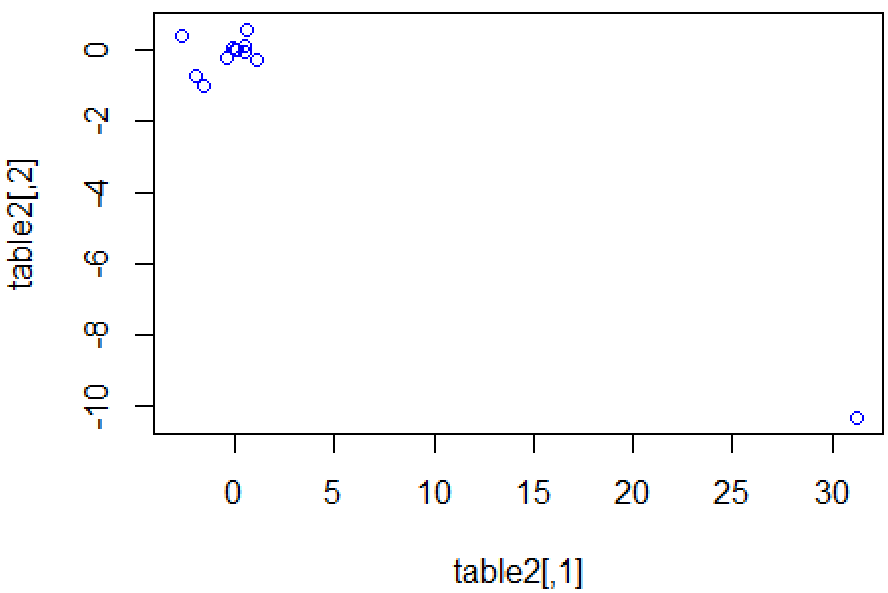
```
summary(lm(crim~.,data=Boston))

##
## Call:
## lm(formula = crim ~ ., data = Boston)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.924 -2.120 -0.353  1.019 75.051
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  17.033228   7.234903   2.354 0.018949 *
## zn            0.044855   0.018734   2.394 0.017025 *
## indus        -0.063855   0.083407  -0.766 0.444294
## chas         -0.749134   1.180147  -0.635 0.525867
## nox         -10.313535   5.275536  -1.955 0.051152 .
## rm            0.430131   0.612830   0.702 0.483089
## age           0.001452   0.017925   0.081 0.935488
## dis          -0.987176   0.281817  -3.503 0.000502 ***
## rad           0.588209   0.088049   6.680 6.46e-11 ***
## tax          -0.003780   0.005156  -0.733 0.463793
## ptratio      -0.271081   0.186450  -1.454 0.146611
## black        -0.007538   0.003673  -2.052 0.040702 *
## lstat         0.126211   0.075725   1.667 0.096208 .
## medv         -0.198887   0.060516  -3.287 0.001087 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.439 on 492 degrees of freedom
## Multiple R-squared:  0.454,  Adjusted R-squared:  0.4396
## F-statistic: 31.47 on 13 and 492 DF,  p-value: < 2.2e-16
```

不只有 chas 沒拒絕$H_0$，還有 indus,nox….

### (c)

```
table2<-cbind(a[,3],lm(crim~.,data=Boston)$coef[-1])
plot(table2,col=12)
```




```
rcorr(as.matrix(Boston))

##            crim    zn indus  chas   nox    rm   age   dis   rad   tax
## crim       1.00 -0.20  0.41 -0.06  0.42 -0.22  0.35 -0.38  0.63  0.58
## zn        -0.20  1.00 -0.53 -0.04 -0.52  0.31 -0.57  0.66 -0.31 -0.31
## indus      0.41 -0.53  1.00  0.06  0.76 -0.39  0.64 -0.71  0.60  0.72
## chas      -0.06 -0.04  0.06  1.00  0.09  0.09  0.09 -0.10 -0.01 -0.04
## nox        0.42 -0.52  0.76  0.09  1.00 -0.30  0.73 -0.77  0.61  0.67
## rm        -0.22  0.31 -0.39  0.09 -0.30  1.00 -0.24  0.21 -0.21 -0.29
## age        0.35 -0.57  0.64  0.09  0.73 -0.24  1.00 -0.75  0.46  0.51
## dis       -0.38  0.66 -0.71 -0.10 -0.77  0.21 -0.75  1.00 -0.49 -0.53
## rad        0.63 -0.31  0.60 -0.01  0.61 -0.21  0.46 -0.49  1.00  0.91
## tax        0.58 -0.31  0.72 -0.04  0.67 -0.29  0.51 -0.53  0.91  1.00
## ptratio    0.29 -0.39  0.38 -0.12  0.19 -0.36  0.26 -0.23  0.46  0.46
## black     -0.39  0.18 -0.36  0.05 -0.38  0.13 -0.27  0.29 -0.44 -0.44
## lstat      0.46 -0.41  0.60 -0.05  0.59 -0.61  0.60 -0.50  0.49  0.54
## medv      -0.39  0.36 -0.48  0.18 -0.43  0.70 -0.38  0.25 -0.38 -0.47
##           ptratio black lstat  medv
## crim         0.29 -0.39  0.46 -0.39
## zn          -0.39  0.18 -0.41  0.36
## indus        0.38 -0.36  0.60 -0.48
## chas        -0.12  0.05 -0.05  0.18
```

```
## nox        0.19 -0.38  0.59 -0.43
## rm        -0.36  0.13 -0.61  0.70
## age        0.26 -0.27  0.60 -0.38
## dis       -0.23  0.29 -0.50  0.25
## rad        0.46 -0.44  0.49 -0.38
## tax        0.46 -0.44  0.54 -0.47
## ptratio    1.00 -0.18  0.37 -0.51
## black     -0.18  1.00 -0.37  0.33
## lstat      0.37 -0.37  1.00 -0.74
## medv      -0.51  0.33 -0.74  1.00
##
## n= 506
##
##
## P
##         crim   zn     indus  chas   nox    rm     age    dis    rad

## crim           0.0000 0.0000 0.2094 0.0000 0.0000 0.0000 0.0000 0.00
00
## zn      0.0000        0.0000 0.3378 0.0000 0.0000 0.0000 0.0000 0.00
00
## indus   0.0000 0.0000        0.1575 0.0000 0.0000 0.0000 0.0000 0.00
00
## chas    0.2094 0.3378 0.1575        0.0403 0.0402 0.0518 0.0257 0.86
87
## nox     0.0000 0.0000 0.0000 0.0403        0.0000 0.0000 0.0000 0.00
00
## rm      0.0000 0.0000 0.0000 0.0402 0.0000        0.0000 0.0000 0.00
00
## age     0.0000 0.0000 0.0000 0.0518 0.0000 0.0000        0.0000 0.00
00
## dis     0.0000 0.0000 0.0000 0.0257 0.0000 0.0000 0.0000        0.00
00
## rad     0.0000 0.0000 0.0000 0.8687 0.0000 0.0000 0.0000 0.0000

## tax     0.0000 0.0000 0.0000 0.4244 0.0000 0.0000 0.0000 0.0000 0.00
00
## ptratio 0.0000 0.0000 0.0000 0.0062 0.0000 0.0000 0.0000 0.0000 0.00
00
## black   0.0000 0.0000 0.0000 0.2733 0.0000 0.0039 0.0000 0.0000 0.00
00
## lstat   0.0000 0.0000 0.0000 0.2259 0.0000 0.0000 0.0000 0.0000 0.00
00
## medv    0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.00
00
##         tax    ptratio black  lstat  medv
## crim    0.0000 0.0000  0.0000 0.0000 0.0000
## zn      0.0000 0.0000  0.0000 0.0000 0.0000
## indus   0.0000 0.0000  0.0000 0.0000 0.0000
## chas    0.4244 0.0062  0.2733 0.2259 0.0000
```

```
## nox     0.0000 0.0000  0.0000 0.0000 0.0000
## rm      0.0000 0.0000  0.0039 0.0000 0.0000
## age     0.0000 0.0000  0.0000 0.0000 0.0000
## dis     0.0000 0.0000  0.0000 0.0000 0.0000
## rad     0.0000 0.0000  0.0000 0.0000 0.0000
## tax            0.0000  0.0000 0.0000 0.0000
## ptratio 0.0000         0.0000 0.0000 0.0000
## black   0.0000 0.0000         0.0000 0.0000
## lstat   0.0000 0.0000  0.0000        0.0000
## medv    0.0000 0.0000  0.0000 0.0000
```

在$H_0: \rho = 0$假設下，只有跟 Chas 不拒絕，其餘皆線性獨立則 crim 若配適 n 個解釋變數模型，且解釋變數都是 boston 的變數，則皆不會有太強烈的線性重和問題，若解釋變數中的相關係數很高，則非常容易線性重和導致模型不夠穩定。

## (d)

```
table2<-Map(summary,(Map(function(x)lm(crim~poly(x,3),data=Boston),Bost
on[,c(-1,-4,-9)])))
a2<-data.frame(11,8)
for(i in 1:11)
{

a2[i,1]<-Map(function(x)lm(crim~poly(x,3),data=Boston),Boston[,c(-1,-4,
-9)])[[i]]$coef[1]

a2[i,2]<-table2[[i]]$coefficients[13]

a2[i,3]<-Map(function(x)lm(crim~poly(x,3),data=Boston),Boston[,c(-1,-4,
-9)])[[i]]$coef[2]

a2[i,4]<-table2[[i]]$coefficients[14]

a2[i,5]<-Map(function(x)lm(crim~poly(x,3),data=Boston),Boston[,c(-1,-4,
-9)])[[i]]$coef[3]

a2[i,6]<-table2[[i]]$coefficients[15]

a2[i,7]<-Map(function(x)lm(crim~poly(x,3),data=Boston),Boston[,c(-1,-4,
-9)])[[i]]$coef[4]

a2[i,8]<-table2[[i]]$coefficients[16]
}
colnames(a2)<-c("b_0","p-value","b_1","p-value","b_2","p-value","b_3")
rownames(a2)<-colnames(Boston)[c(-1,-4,-9)]
a2

##              b_0      p-value       b_1      p-value        b_2
## zn      3.613524 1.547150e-20 -38.74984 4.697806e-06  23.939832
## indus   3.613524 3.606468e-25  78.59082 8.854243e-24 -24.394796
```

```
## nox     3.613524 2.742908e-26  81.37202 2.457491e-26 -28.828594
## rm      3.613524 1.026665e-20 -42.37944 5.128048e-07  26.576770
## age     3.613524 5.918933e-23  68.18201 4.878803e-17  37.484470
## dis     3.613524 1.060226e-25 -73.38859 1.253249e-21  56.373036
## tax     3.613524 8.955923e-29 112.64583 6.976314e-49  32.087251
## ptratio 3.613524 1.270767e-21  56.04523 1.565484e-11  24.774824
## black   3.613524 2.139710e-22 -74.43120 2.730082e-19   5.926419
## lstat   3.613524 4.939398e-24  88.06967 1.678072e-27  15.888164
## medv    3.613524 7.024110e-31 -75.05761 4.930818e-27  88.086211
##              p-value        b_3           NA
## zn      4.420507e-03 -10.071868 2.295386e-01
## indus   1.086057e-03 -54.129763 1.196405e-12
## nox     7.736755e-05 -60.361894 6.961110e-16
## rm      1.508545e-03  -5.510342 5.085751e-01
## age     2.291156e-06  21.353207 6.679915e-03
## dis     7.869767e-14 -42.621877 1.088832e-08
## tax     3.665348e-06  -7.996811 2.438507e-01
## ptratio 2.405468e-03 -22.279737 6.300514e-03
## black   4.566044e-01  -4.834565 5.436172e-01
## lstat   3.780418e-02 -11.574022 1.298906e-01
## medv    2.928577e-35 -48.033435 1.046510e-12
```

只有 indus、dis、nox 、age、ptratio、medv 有 1、2、3 次相關 顯著

1. Using a little bit of algebra, prove that (4.2) is equivalent to (4.3). In other words, the logistic function representation and logit representation for the logistic regression model are equivalent.

2. It was stated in the text that classifying an observation to the class for which (4.12) is largest is equivalent to classifying an observation to the class for which (4.13) is largest. Prove that this is the case. In other words, under the assumption that the observations in the $k$th class are drawn from a $N(\mu_k, \sigma^2)$ distribution, the Bayes' classifier assigns an observation to the class for which the discriminant function is maximized.

3. This problem relates to the QDA model, in which the observations within each class are drawn from a normal distribution with a class-specific mean vector and a class specific covariance matrix. We consider the simple case where $p = 1$; i.e. there is only one feature.

   Suppose that we have $K$ classes, and that if an observation belongs to the $k$th class then $X$ comes from a one-dimensional normal distribution, $X \sim N(\mu_k, \sigma_k^2)$. Recall that the density function for the one-dimensional normal distribution is given in (4.11). Prove that in this case, the Bayes' classifier is *not* linear. Argue that it is in fact quadratic.

   *Hint: For this problem, you should follow the arguments laid out in Section 4.4.2, but without making the assumption that $\sigma_1^2 = \ldots = \sigma_K^2$.*

## (1)

$$\mathbf{P}(x) = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}} = 1 - \frac{1}{1+e^{\beta_0+\beta_1 x}}$$

$$\frac{1}{1-\mathbf{P}(x)} = 1 + e^{\beta_0+\beta_1 x}$$

$$\frac{\mathbf{P}(x)}{1-\mathbf{P}(x)} = e^{\beta_0+\beta_1 x}$$

## (2)

$$p_k(x) = \frac{\pi_k \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{(x-\mu_k)^2}{2\sigma^2})}{\sum_{k=1}^{K} \pi_k \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{(x_k-\mu_k)^2}{2\sigma^2})}$$

$$\max_k p_k(x) \Rightarrow \max_k \pi_k \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{(x-\mu_k)^2}{2\sigma^2}) \Rightarrow \min_k - \log\pi_k + \frac{(x-\mu_k)^2}{2\sigma^2}$$

$$\Rightarrow \min_k x \times \frac{\mu_k}{\sigma} + \frac{\mu_k^2}{2\sigma^2} - \log\pi_k$$

## (3)

$$p_k(x) = \frac{\pi_k f(\mathbf{x})}{\sum_{k=1}^{K} \pi_k f(\mathbf{x})}$$

$$\max_k p_k(x) \Rightarrow \max_k \pi_k f(\mathbf{x}) \Rightarrow \max_k \log\pi_k + \log f(\mathbf{x})$$

$$\Rightarrow \min_k \; - \log\pi_k + \frac{1}{2}(\mathbf{x}-\mu)' \Sigma_k{}^{-1}(\mathbf{x}-\mu) + \frac{1}{2}\log|\Sigma_k|$$