APLM Final exam 610611105 黃三騰

(1)(20 pts) Please explain what the ridge and Lasso regression are, and the difference between the least square estimator, the ridge estimator and Lasso estimator.

Sol:

Based on linear model:

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \epsilon_i \qquad , i = 1, \ldots, n$$

Ridge regression:

$$\min_{\beta} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2 \quad subject\ to\ ||\beta||_2^2 \le t\ , t > 0$$

$$\Leftrightarrow \quad \min_{\beta} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \quad , \lambda > 0$$

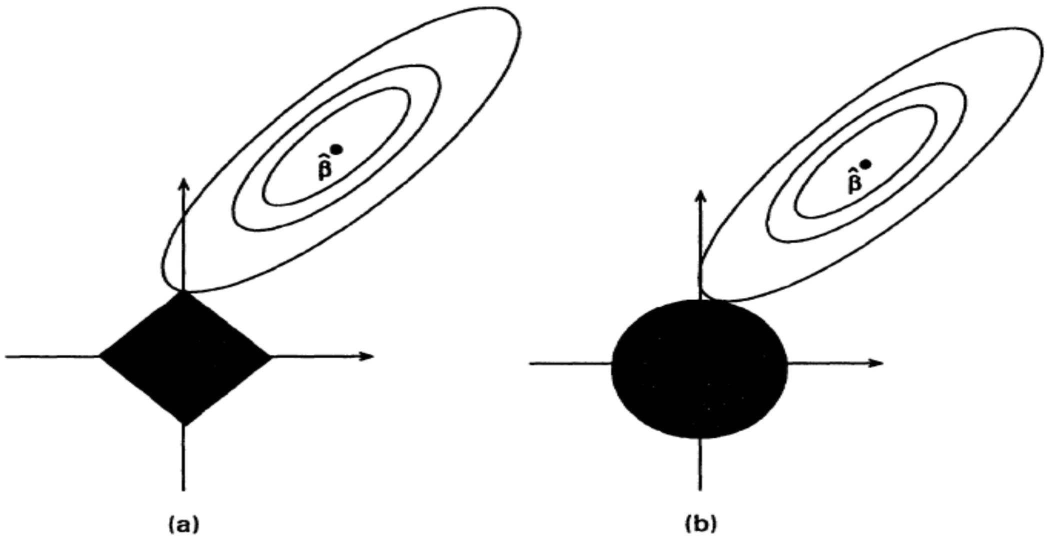
We have Ridge regression estimator is $\hat{\beta}_{ridge} = (X^T X + \lambda I)^{-1} X^T Y$, which is not unbiased(i.e. $E(\hat{\beta}_{ridge}) \neq \beta$).

LASSO:

$$\min_{\beta} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2 \quad subject\ to\ ||\beta||_1 \le t\ , t > 0$$

$$\Leftrightarrow \quad \min_{\beta} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \quad , \lambda > 0$$

	Unbiased	Shrinkage coefficients	Sparsity
Ordinary Least Squares	○	×	×
Ridge	×	○	×
LASSO	×	○	○



Estimation picture for (a) the lasso and (b) ridge regression

Because of the L1-norm, the LASSO can set coefficients exactly equal to zero while Ridge and Ordinary Least Squares can not.

However, the LASSO is not unbiased since the penalty function is $p_{\lambda}(|\beta|) = \lambda|\beta|$. It cause that the larger coefficients is, the more penalized value would get.

(3) A manager would like to increase the return on an investment, which depends on 17 financial products. But, some of financial products can not affect the return. In the attached data set including 2000 data, the first 17 columns show the investment for each product, and the last one is the corresponding return.

(a) (25 pts) Please analyze the data (It is required to provide ANOVA table.)?

(b) Assume that the budget (US 100000) is limited. Based on your analysis results in (a) and (b),

1. (15 pts) please make your comparison on two different strategies (20000,30000,0,40000,10000,0....,0) and (50000,0,20000,0,30000,0....,0).
2. (20 pts) Also, please discuss the prediction of their return throught the theoretical way.

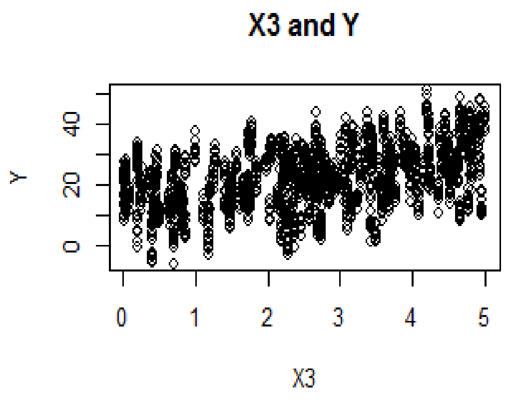
sol:

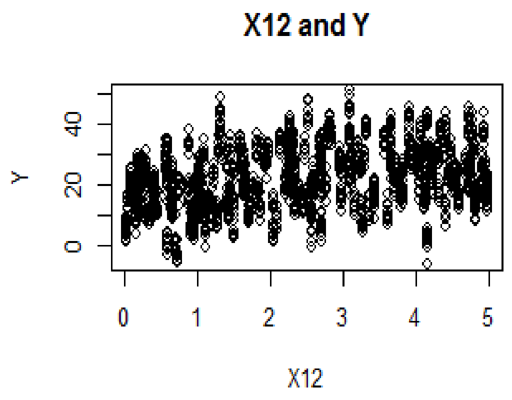
(a)

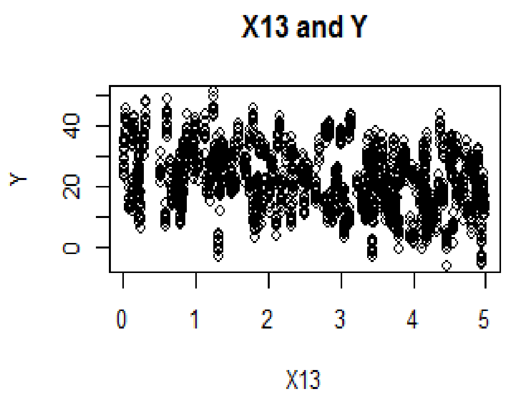
(i) Scatter plot of each Predictor($X_1, \ldots, X_{17}$) and Response(Y).

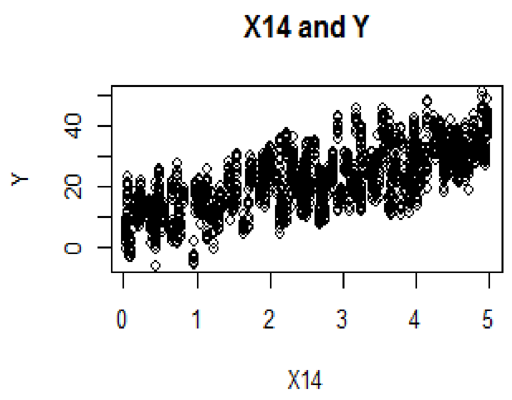
將所有解釋變數($X_1, \ldots, X_{17}$)對反應變數(Y)的散布圖畫出，列出有明顯線性關係的變數:

分別是 $X_3, X_{12}, X_{13}, X_{14}$。(其它解釋變數的圖則無明顯的線性趨勢)









初步判斷後，決定使用線性模型分析。

備註: 這裡我們不考慮將截距項β_0放入模型中，原因為我們的資料是有關 17 種金融產品投資金額($X_1, \dots, X_{17}$)與其投資後的報酬(Y)，而如果沒有投資金額便不會有投資後的報酬。同時也可以看到上面的散布圖是通過原點的，故截距項β_0應為 0。

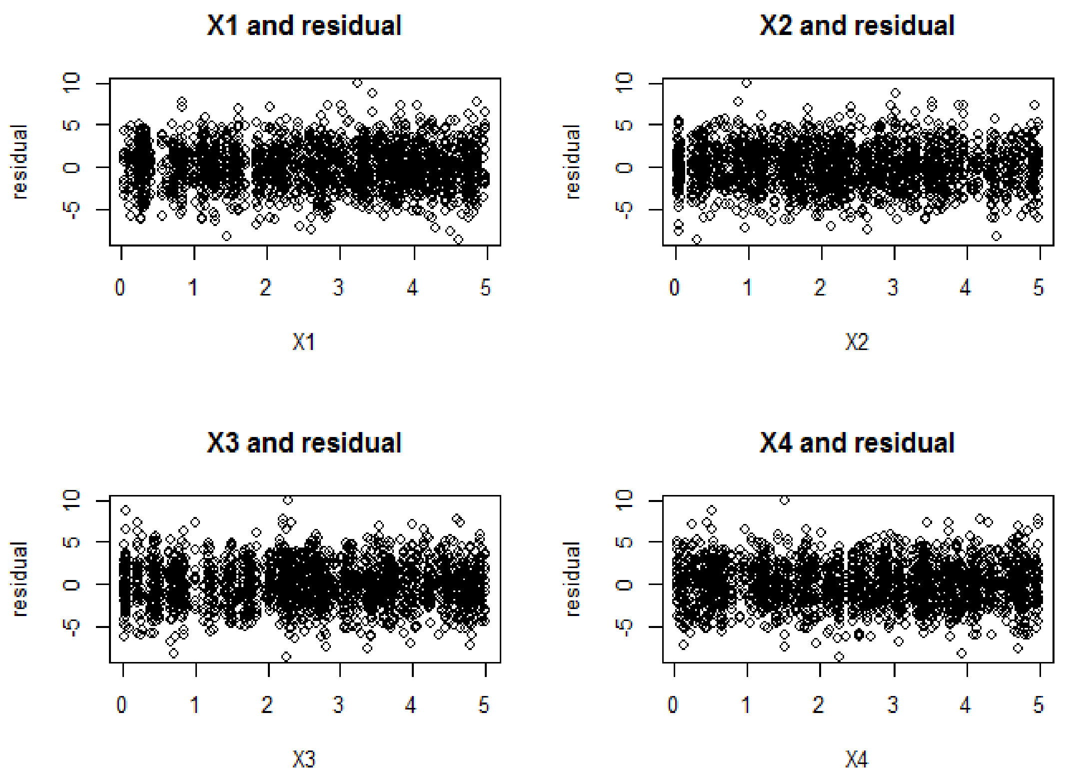
(ii)

Consider the model:

$$Y_i = \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_{17} X_{17} + \epsilon_i \ \ , i = 1, \dots, 2000$$

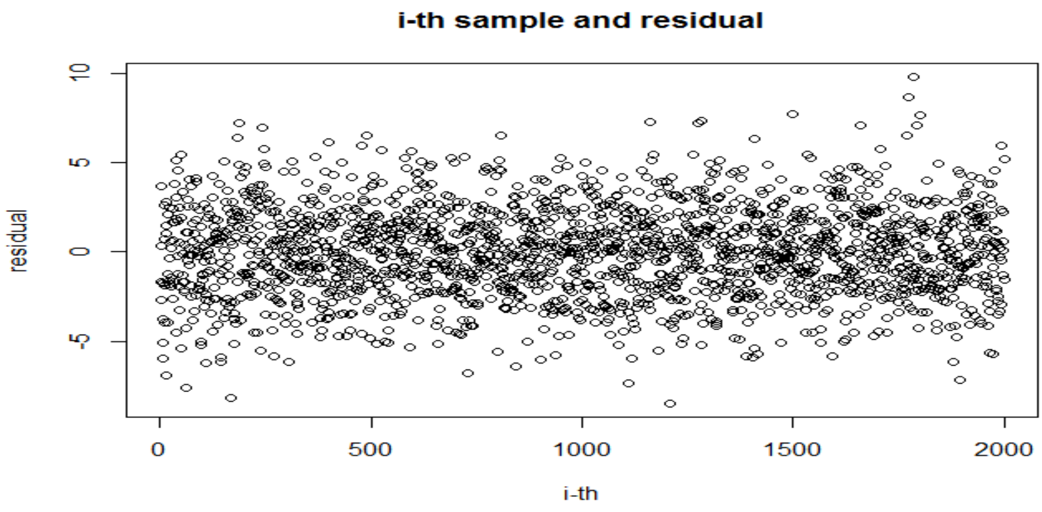
A. 建立線性模型後，我們先來觀察殘差項(e_i)的畫圖:

   1. 解釋變數($X_1, \dots, X_{17}$)與殘差項(e_i)



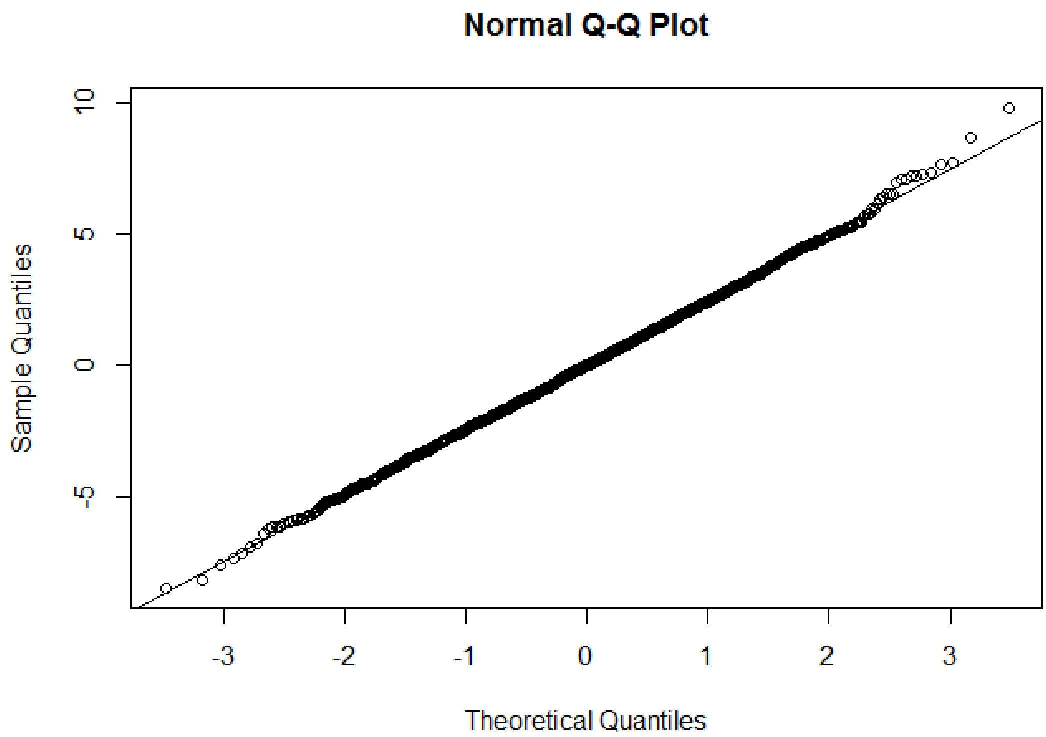
解釋變數與殘差呈現水平無特別模式的關係。(解釋變數X_5~X_{17}的結果與上圖相似，故此處只列出前四個解釋變數)

   2. 資料的順序(時間先後)與殘差項(e_i)



呈現水平無特別模式的關係，代表沒有違背ϵ的變異數σ^2為常數的假設，以及模型並不需考慮將資料的順序(時間先後)納入解釋變數。

3. 殘差項(e_i)的 Q-Q Plot



代表沒有違背ϵ服從常態分配的假設。

B. 使用線性模型估計的結果:

Residuals:				
Min	1Q	Median	3Q	Max
-8.5003	-1.6910	0.0214	1.6751	9.8326

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
X1	-0.4903032	0.0391813	-12.514	< 2e-16 ***
X2	-0.0216555	0.0413625	-0.524	0.6006
X3	3.0576619	0.0382537	79.931	< 2e-16 ***
X4	0.0086003	0.0378851	0.227	0.8204
X5	1.1838885	0.0399983	29.598	< 2e-16 ***
X6	0.1387319	0.0351258	3.950	8.1e-05 ***
X7	0.0812278	0.0418093	1.943	0.0522 .
X8	-0.0066523	0.0384489	-0.173	0.8627
X9	0.0006042	0.0429780	0.014	0.9888
X10	-0.0296778	0.0408369	-0.727	0.4675
X11	0.0583555	0.0374444	1.558	0.1193
X12	2.0379867	0.0363056	56.134	< 2e-16 ***
X13	-2.4571694	0.0383099	-64.139	< 2e-16 ***
X14	4.9831330	0.0399680	124.678	< 2e-16 ***
X15	0.0876630	0.0370593	2.365	0.0181 *
X16	-0.0096647	0.0408550	-0.237	0.8130
X17	0.0285892	0.0357266	0.800	0.4237

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.49 on 1983 degrees of freedom

Multiple R-squared: 0.9893, Adjusted R-squared: 0.9892

F-statistic: 1.079e+04 on 17 and 1983 DF, p-value: < 2.2e-16

此表提供$\beta_1, \beta_2, \ldots, \beta_{17}$的估計值，以及

各係數是否為 0 之假設檢定: $\begin{matrix} H_0: \beta_j = 0 \\ H_a: \beta_j \neq 0 \end{matrix}$ $, j = 1, \ldots, 17$.

這裡$R^2 = 0.9893$和$R^2_{\text{adjust}} = 0.9892$都很高，代表解釋變數$(X_1, \ldots, X_{17})$與反應變數(Y)是有很強的線性關係。

此外，

ANOVA Table

Source	df	Sum Sq	Mean Sq	F	P-value
Regression	16	1136975	71060.96	11462.18	< 2.2e-16
Error	1983	12294	6.1996		
Total	1999				

ANOVA 表給出解釋變數(X_j)與反應變數(Y): $\begin{matrix} H_0:\ \beta_1 = \beta_2 = \cdots = \beta_{17} = 0 \\ H_a: some\ of\ \beta_j \neq 0\ , j = 1, \dots, 17 \end{matrix}$的假設檢定結果，解釋變數($X_1, \dots, X_{17}$)與反應變數(Y)是有線性關係的。

(iii)
由(ii)的分析結果，得到解釋變數($X_1, \dots, X_{17}$)與反應變數(Y)是有線性關係。然而，在一些解釋變數上的$\begin{matrix} H_0: \beta_j = 0 \\ H_a: \beta_j \neq 0 \end{matrix}$ 假設檢定顯示無法拒絕H_0，亦即模型中可能有些是不重要的解釋變數($\beta_j = 0, j = 1, \dots, 17$)。
這裡我們打算利用 Forward Stepwise regression 的方式，進行選模:
(備註: 操作的細節是參照上課筆記，設定$\alpha_{\text{enter}} = 0.01, \alpha_{drop} = 0.05$)

以下表格紀錄選模過程，

<u>Step 0</u>
空模型.
<u>Step 1</u>
X_{14}加入模型中，此時: $\{X_{14}\}$.
<u>Step 2 and Step 3</u>
X_3加入模型中，此時: $\{X_{14}, X_3\}$ ；沒有變數被刪除.
X_{12}加入模型中，此時: $\{X_{14}, X_3, X_{12}\}$ ；沒有變數被刪除.
X_{13}加入模型中，此時: $\{X_{14}, X_3, X_{12}, X_{13}\}$ ；沒有變數被刪除.
X_5加入模型中，此時: $\{X_{14}, X_3, X_{12}, X_{13}, X_5\}$ ；沒有變數被刪除.
X_1加入模型中，此時: $\{X_{14}, X_3, X_{12}, X_{13}, X_5, X_1\}$ ；沒有變數被刪除.
X_6加入模型中，此時: $\{X_{14}, X_3, X_{12}, X_{13}, X_5, X_1, X_6\}$ ；沒有變數被刪除.
沒有變數加入，且沒有變數被刪除，最後模型為: $\{X_{14}, X_3, X_{12}, X_{13}, X_5, X_1, X_6\}$

利用選模結果，$\{Y_i, X_{i1}, X_{i3}, X_{i5}, X_{i6}, X_{i12}, X_{i13}, X_{i14}\}_{i=1}^{2000}$建立線性模型:

$$Y_i = \beta_1 X_{i1} + \beta_3 X_{i3} + \beta_5 X_{i5} + \beta_6 X_{i6} + \beta_{i12} X_{12} + \beta_{i13} X_{13} + \beta_{14} X_{i14} + \epsilon_i\ ,$$
$$i = 1, \dots, 2000$$

Residuals:

Min	1Q	Median	3Q	Max
-8.4709	-1.6916	0.0339	1.6861	10.0568

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
X1	-0.44335	0.03504	-12.652	< 2e-16 ***
X3	3.08859	0.03578	86.330	< 2e-16 ***

```
X5     1.19762      0.03731    32.096    < 2e-16 ***
X6     0.14796      0.03306     4.476    8.04e-06 ***
X12    2.06026      0.03403    60.541    < 2e-16 ***
X13    -2.42925     0.03503    -69.348   < 2e-16 ***
X14    5.01280      0.03818    131.308   < 2e-16 ***
---
Signif. codes:   0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.492 on 1993 degrees of freedom
Multiple R-squared:   0.9892,      Adjusted R-squared:   0.9892
F-statistic: 2.615e+04 on 7 and 1993 DF,   p-value: < 2.2e-16
```

得到估計值，

$$(\hat{\beta}_1, \hat{\beta}_3, \hat{\beta}_5, \hat{\beta}_6, \hat{\beta}_{12}, \hat{\beta}_{13}, \hat{\beta}_{14}) = (-0.44\,, 3.09\,, 1.2\,, 0.15\,, 2.06\,, -2.43\,, 5.01)$$

(b)

1.

根據兩種投資策略:

strage1 $(\mathrm{X}_1, \mathrm{X}_2, \mathrm{X}_3, \mathrm{X}_4, \mathrm{X}_5, \dots, \mathrm{X}_{17}) = (20000\,, 30000\,, 0\,, 40000\,, 10000\,, 0, \dots, 0)$

strage2 $(\mathrm{X}_1, \mathrm{X}_2, \mathrm{X}_3, \mathrm{X}_4, \mathrm{X}_5, \dots, \mathrm{X}_{17}) = (50000\,, 0\,, 20000, 0, 30000, 0, \dots, 0)$

利用(a)最終模型的估計值，

$$(\hat{\beta}_1, \hat{\beta}_3, \hat{\beta}_5, \hat{\beta}_6, \hat{\beta}_{12}, \hat{\beta}_{13}, \hat{\beta}_{14}) = (-0.44\,, 3.09\,, 1.2\,, 0.15\,, 2.06\,, -2.43\,, 5.01)$$

求得

$$\hat{Y}_{strage1} = 3109.167\ \ , \hat{Y}_{strage2} = 75532.766$$

我們發現 strage2 的投資方式是可以得到比較多的報酬。

主要原因為: 同樣預算下，strage2 投資在$(\mathrm{X}_1, \mathrm{X}_3, \mathrm{X}_5) = (50000, 20000, 30000)$；而 strage1 投資在$(\mathrm{X}_1, \mathrm{X}_3, \mathrm{X}_5) = (20000, 0, 10000)$。根據我們的模型，投資在$(\mathrm{X}_3, \mathrm{X}_5)$能夠得到報酬，而投資在$\mathrm{X}_1$則會產生虧損。

2.

Given a new data, $\mathbf{x}_{\text{new}} = (x_1, x_2, \dots, x_p)_{1\times p}$, the corresponding response Y_{new}.

To construct the 95% confidence interval for Y_{new},

$\Rightarrow \mathrm{Y}_{\text{new}} \sim N(\mathbf{x}_{\text{new}}\boldsymbol{\beta}, \sigma^2)$, $and\ \mathrm{Y}_{\text{new}}\ independent\ with\ \hat{Y}_{new} = \mathbf{x}_{\text{new}}\hat{\boldsymbol{\beta}}$

$\Rightarrow \hat{Y}_{new} - \mathrm{Y}_{\text{new}} \sim N(0, \sigma^2 + \sigma^2 \mathbf{x}_{\text{new}}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\mathbf{x}_{\text{new}}^T)$

Replace σ^2 by MSE,

$$\frac{\hat{Y}_{new} - \mathrm{Y}_{\text{new}}}{\sqrt{MSE(1 + \mathbf{x}_{\text{new}}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\mathbf{x}_{\text{new}}^T)}} \sim \mathrm{t}_{\mathrm{n-p}}$$

Hence, $(1-\alpha)\%$ confidence interval for Y_{new} is

$$[\ \hat{Y}_{new}\ \pm\ t_{\frac{\alpha}{2},n-p}\sqrt{MSE(1+\mathbf{x}_{\text{new}}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\mathbf{x}_{\text{new}}^T)}\ ]$$

Here,

95% confidence interval for Y_{strage1} is $[\,1735.42\ ,\ 4482.92\,]$ and

95% confidence interval for Y_{strage2} is $[\,72088.17\ ,\ 78977.36\,]$.

從信賴區間我們得到兩種投資策略的報酬所落在的範圍，故判斷 strage2 的投資方式可以得到比 strage1 多的報酬。

<u>附錄:</u>

在(3)分析資料時，我同時也嘗試了 Penalized Least Squares 的方式進行選模:

我們使用 LASSO、SCAD 的方式估計，λ利用 10-folds Cross Validation 選出:

方法	選到的模型(X_j以所對應的$\lvert\beta_j\rvert$值大到小列出)
Forward Stepwise	$\{X_{14},X_3,X_{13},X_{12},X_5,X_1,X_6\}$
LASSO($\lambda=0.088$)	$\{X_{14},X_3,X_{13},X_{12},X_5,X_1,X_6,X_{15},X_7,X_{11},X_{17}\}$
SCAD($\lambda=0.162$)	$\{X_{14},X_3,X_{13},X_{12},X_5,X_1,X_6,X_{15},X_7,X_{11}\}$

方法	估計值 $\hat{\beta}_j$, $j\in\{1,\ldots,17\}$
Forward Stepwise	$(\hat{\beta}_{14},\hat{\beta}_3,\hat{\beta}_{13},\hat{\beta}_{12},\hat{\beta}_5,\hat{\beta}_1,\hat{\beta}_6)=$ $(5.01,3.09,-2.43,2.06,1.2,-0.44,0.15)$
LASSO($\lambda=0.088$)	$(\hat{\beta}_{14},\hat{\beta}_3,\hat{\beta}_{13},\hat{\beta}_{12},\hat{\beta}_5,\hat{\beta}_1,\hat{\beta}_6,\hat{\beta}_{15},\hat{\beta}_7,\hat{\beta}_{11},\hat{\beta}_{17})=$ $(4.97,3.04,-2.39,2.02,1.15,-0.41,0.12,0.06,0.03,0.02,0.006)$
SCAD($\lambda=0.162$)	$(\hat{\beta}_{14},\hat{\beta}_3,\hat{\beta}_{13},\hat{\beta}_{12},\hat{\beta}_5,\hat{\beta}_1,\hat{\beta}_6,\hat{\beta}_{15},\hat{\beta}_7,\hat{\beta}_{11})=$ $(5.00,3.08,-2.44,2.05,1.19,-0.45,0.11,0.05,0.03,0.02)$

在這裡 LASSO 與 SCAD 都選到比較大的模型(Forward $\subset$ SCAD $\subset$ LASSO)，推測可能是λ不夠大，無法將係數估計壓到零。(註: 在 LASSO 與 SCAD 的估計結果若$\lvert\beta_j\rvert<10^{-3}$則將它設為零)