統計模擬期末報告

組員：金碩一 107352028 彭昱齊

統碩一 107354014 粘明揚

統碩一 107354015 蔡廷儀

統碩一 107354016 蔡詠丞

Part 1

一、資料簡介

名稱 : Longley

變數個數 : 7 個

資料個數 : 16 筆

簡介 : 本次所用資料為“ Longley”，是一筆在總體經濟學中共線性迴歸有名的範例。

變數介紹 :

GNP.deflator	GNP 縮減指數(1954 年為 100)
GNP	Gross National Product(國民生產毛額)
Unemployed	失業人口
Armed.Forced	從軍人口
Population	大於 14 歲的”適合工作人口”
Year	年度
Employed	就業人口

*其中我們定義一個新變數 real GNP =GNP/GNP.deflator

二、模型介紹

梯度下降法（Gradient descent，GD）

在最佳化中，以一階微分找最佳解的方法，主要是用梯度下降法找到函數的局部最小 值，因為梯度的方向是走向局部最大的方向，所以在梯度下降法中是往梯度的反方向走。

符號介紹：

一維度的純量 x 的梯度，通常用 f ‘ (x) 表示。

多維度的向量 x 的梯度，通常用 ∇ f(x) 表示。

大致作法可分成以下幾步驟：

1. 找目標函數 f(x) 的局部最小值(local minimum)
2. 設立起始值 $\boldsymbol{x_0}$
3. 第 n 次迭代時，將 $\boldsymbol{X_n}$ 往梯度的反方向移動，亦即 $-\nabla Q$

即可表為下式:

$$\beta \leftarrow \beta - r\nabla \sum_{i=1}^{n} Q_i$$

其中 $\boldsymbol{r_0}$ 為學習率 (Learning Rate)

學習率過大，函數可能無法收斂至局部最小值，

學習率過小，函數收斂速度會過於緩慢而難進展。

在梯度下降法中，每次更新參數時，所有的樣本皆有貢獻，所得到的最佳化參數也最為準確，但如果樣本數過大，會造成運算時間過長，為避免此情況，則可考慮採用以下方法。

隨機梯度下降法（Stochastic gradient descent，SGD）

與 GD 不同的是，每次更新參數時僅用 1 個樣本來調整，而隨機的部分即為用 1 個樣本來近似所有樣本調整參數，雖然不是每次迭代都是往最佳化方向移動，但整體而言確實是往最佳化方向移動的，對比於 GD 來說，雖然相對較為不準確，但會更快收斂。

大致作法可分成以下幾步驟：

1. 隨機置換 n 個觀測值
2. 設立起始值 x_0
3. 更新參數直至收斂

即可表為下式:

for i in 1:n, do

$$\boldsymbol{\beta} \leftarrow \boldsymbol{\beta} - r\nabla Q_i$$

repeat above steps until convergence

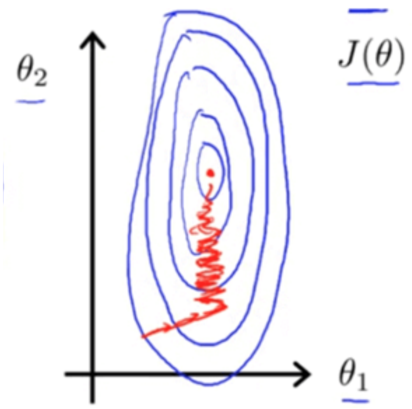
三、SGD 預處理

在使用 SGD 前，要先標準化 longley（反應變數）

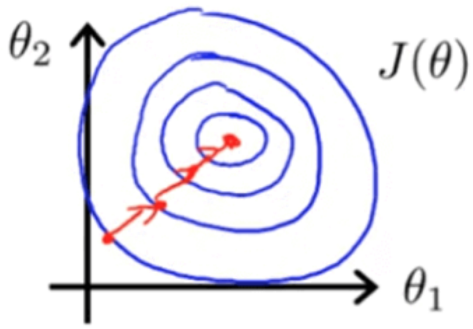
標準化: 即將資料轉換為平均值為 0，標準差為 1 的新資料。

可表為下式:

$$x^* = \frac{x-\mu}{\sigma}$$



<左圖為標準化前的的損失函數>



<右圖為標準化後的損失函數>

未標準化的損失函數 → 各變量的大小範圍大 →等高線圖為橢圓

→ 迭代路徑曲折 → 收斂較慢

已標準化的損失函數 → 各變量的大小範圍相似 → 等高線圖接近正圓

→ 迭代路徑平穩 → 收斂較快

所以在將 longley（反應變數）代入 SGD 模型前，需先進行標準化，再套入 SGD 去找局部最小值。

四、損失函數 cost function

目標：極小化損失函數。

透過 stochastic gradient decent 迭代，每一次的迭代都隨機從 16 筆資料中抽取一筆資料進行參數更新，設定 learning rate（r）（此資料設定 r = 0.01），直到$\mathbf{step\ size} = \mathbf{r} * \nabla\mathbf{Q} < \boldsymbol{\varepsilon}$（此題設為$\mathbf{10^{-4}}$）停止迭代。

$$遞迴式：\beta_j{}^{(t+1)} = \beta_j{}^{(t)} - \mathrm{r} * \nabla Q_i{}^{(t)}$$

Linear Regression

$$Cost\ function = \sum_{i=1}^{N}\left\{y_i - \sum_{j=0}^{p}\beta_j x_{ij}\right\}^2 = (\mathrm{y}-\mathrm{X}\beta)^T(\mathrm{y}-\mathrm{X}\beta) + \lambda\beta^T\beta$$

$$\nabla Q_i = (-2)\sum_{i=1}^{N}\left\{y_i - \sum_{j=0}^{p}\beta_j x_{ij}\right\}x_{ij} = X^T(X^T\beta - \mathrm{y})$$

Ridge Regression

Ridge Regression 透過將懲罰參數 $\boldsymbol{\lambda\sum_{j=1}^{p}\beta_j{}^2}$ 加入目標函式中，也因該參數為對係數做出二階懲罰，故又稱 L2 Penalty 懲罰參數。

此時我們透過 cross-validation(3-fold)，去找一最佳$\boldsymbol{\lambda} = \mathbf{0.09189027}$來進行參數的迭代。

$$Cost\ function = \sum_{i=1}^{N}\left\{y_i - \sum_{j=0}^{p}\beta_j x_{ij}\right\}^2 + \lambda\sum_{j=1}^{p}\beta_j{}^2 = (\mathrm{y}-\mathrm{X}\beta)^T(\mathrm{y}-\mathrm{X}\beta) + \lambda\beta^T\beta$$

$$\nabla Q_i = (-2)\sum_{i=1}^{N}\left\{y_i - \sum_{j=0}^{p}\beta_j x_{ij}\right\}x_{ij} + 2\lambda\sum_{j=1}^{p}\beta_j = X^T(X^T\beta - y) + 2\lambda\beta$$

Advantages & Disadvantages

優點

1. 會將具有相關性的變數推向彼此，避免產生兩變數的係數有一極大正、一極大負的情形。

2. 許多不相干的變數係數會被逼近為 0(不會等於 0)。表示我們可以降低我們資料中的雜訊，更清楚的識別出模型中真正的訊號(signals)。

缺點

1. 會保留所有變數。適用情況為需要保留所有變數並將較無影響力的變數雜訊給減弱並最小化共線性。
2. 不具變數挑選(feature selection)功能，如需挑選變數，則使用 lasso 較佳。

五、資料分析

首先由於 GNP.deflator 及 GNP 兩個變數間存在共線性問題，所以以 realGNP 來取代這 2 個變數，然後再分別對這筆資料以 SGD 尋找 LSE 及 Ridge Regression 的參數，並與其理論值做比較，結果如下:

Learning rate= 0.01

停止迭代條件：x_t-$x_{t\text{-}1} \leq 10^{-6}$

或迭代次數大於10^7次

	起始值	*SGD for LSE*	*GD for LSE*	*LSE 理論值*
Intercept	0	*-0.007279764*	*-1.831564e-15*	*-1.605469e-15*
Unemployed	1	*-0.341234635*	*-4.536119e-01*	*-4.691494e-01*
Year	1	*2.575378720*	*2.406119e+00*	*2.602258e+00*
Armed.Forces	*1*	*-1.450544729*	*-1.586930e-01*	*-1.533544e-01*
Population	*1*	*-0.648098149*	*-6.431348e-01*	*-7.160525e-01*
realGNP	*1*	*-0.458285761*	*-4.325817e-01*	*-5.496801e-01*
迭代次數		*332*	*470961*	

Learning rate= $\mathbf{0.001}$

停止迭代條件：$\boldsymbol{x_t}$-$\boldsymbol{x_{t-1}} \leq \mathbf{10^{-6}}$

或迭代次數大於$\mathbf{10^6}$次

	起始值	*SGD for Ridge*	*GD for Ridge*	*Ridge 理論值*
Intercept	*0*	*-1.272181e-04*	*-7.761552e-17*	*-1.373901e-15*
Unemployed	*-0.3*	*-8.618608e-05*	*1.729801e-02*	*-3.083591e-01*
Year	*-0.1*	*-3.043305e-05*	*3.482318e-02*	*-1.765277e-01*
Armed.Forces	*0.6*	*7.241449e-04*	*1.632026e-02*	*6.023633e-01*
Population	*0.1*	*3.059301e-04*	*3.442245e-02*	*1.392591e-01*
realGNP	*0.5*	*6.409010e-04*	*3.522200e-02*	*5.140326e-01*
迭代次數		*35*	*46*	

六、資料解讀

分別用 GD 及 SGD 的模型下，GD 較靠近理論值，迭代次數也較多，兩者方法各有好處，我們可以依據資料量或其他因素來選擇方法。

資料量較大，考慮 SGD 來減少資料儲存容量，加快迭代速度

資料量較少，考慮 GD 來處理，結果準確，耗費時間及記憶體也不會消耗太多

我們發現 LSE 及 Ridge 的參數明顯不同，甚至產生變號情形，猜測可能是因為 Ridge 有懲罰項，原本差異大的係數，被強迫靠近彼此導致收斂方向不同，所以 Ridge 的參數係數也不會產生相差太大的情況。

Part 2

Option pricing (improve Monte Carlo simulation in homework)

反向變異法(Antithetic Method)

在隨機抽樣中，將抽樣到的亂數，取它的相反符號，各得到一正一負的亂數，讓抽樣的均數為 0。假設抽出的亂數為 $\boldsymbol{\varepsilon_1}$、$\boldsymbol{\varepsilon_2}$、⋯、$\boldsymbol{\varepsilon_m}$，算出的選擇權價值為 $\boldsymbol{C_1}$ 和以亂數為 $-\boldsymbol{\varepsilon_1}$、$-\boldsymbol{\varepsilon_2}$、⋯、$-\boldsymbol{\varepsilon_m}$ 算出的選擇權價值為 $\boldsymbol{C_2}$，取平均為 $\boldsymbol{C=\frac{C_1+C_2}{2}}$

$$\boldsymbol{Var(C)=\frac{1}{4}\left[Var(C_1)+Var(C_2)+2Cov(C_1,C_2)\right]}$$

控制變異法(Control Variate Method)

找出與想要模擬的相近的性質或價值且有公式解的選擇權，作為控制變異選擇權。假設 $\boldsymbol{\widehat{C}^a}$ 是模擬出來的算術平均的亞式買權價值，$\boldsymbol{\widehat{C}^g}$ 是模擬出來的幾何平均亞式買權價值，而 $\boldsymbol{C^g}$ 是幾何平均的亞式買權的公式解：

For a geometric average call,

$$
\begin{aligned}
\text{option value} &= S_0 e^{(a-r)T} N(d_1) - K e^{-rT} N(d_2) \\
&= e^{-rT}[S_0 e^{aT} N(d_1) - K N(d_2)] \\
&= e^{-rT}[E[\text{geometric average until } T] N(d_1) - K N(d_2)]
\end{aligned}
$$

$$
\begin{cases}
d_1 = & \frac{\ln(S_0 e^{aT}/K) + (\frac{1}{2}\sigma_G^2)T}{\sigma_G\sqrt{T}} = \frac{\ln(S_0/K) + (a+\frac{1}{2}\sigma_G^2)T}{\sigma_G\sqrt{T}} \\
d_2 = & d_1 - \sigma_G\sqrt{T}
\end{cases}
$$

$$
\begin{cases}
a = & \frac{1}{2}(r - q - \frac{\sigma^2}{6}) \\
\sigma_G = & \frac{\sigma}{\sqrt{3}}
\end{cases}
$$

因 $\boldsymbol{C^a = C^g + E(\hat{C}^a - \hat{C}^g)}$

所以可以假設模擬的算術平均式買權價值為 $\boldsymbol{\hat{C}^a_{cv} = \hat{C}^a + (C^g - \hat{C}^g)}$

因 $\boldsymbol{E(\hat{C}^a_{cv}) = E(\hat{C}^a) + E(C^g - \hat{C}^g) = C^a}$

所以此估計式具不偏性，且因

$$\boldsymbol{Var(\hat{C}^a_{cv}) = Var(\hat{C}^a) + Var(\hat{C}^g) - 2Cov(\hat{C}^a, \hat{C}^g)}$$

若模擬出來的算術平均的亞式買權價值與模擬出來的幾何平均亞式買權價值正相關，則可以降低估計式的變異數。

考慮相關係數的控制變異法

若模擬的算術平均式買權價值設為 $\boldsymbol{\hat{C}^a_{cv} = \hat{C}^a + \beta(C^g - \hat{C}^g)}$

因 $\boldsymbol{E(\hat{C}^a_{cv}) = E(\hat{C}^a) + \beta E(C^g - \hat{C}^g) = C^a}$ 此估計式仍具不偏性

因 $\boldsymbol{Var(\hat{C}^a_{cv}) = Var(\hat{C}^a) + \beta^2 Var(\hat{C}^g) - 2\beta Cov(\hat{C}^a, \hat{C}^g)}$

讓變異數最低的 $\boldsymbol{\beta}$ 為 $\boldsymbol{\beta = \frac{Cov(\hat{C}^a, \hat{C}^g)}{Var(\hat{C}^g)}}$

所以當模擬出來的算術平均的亞式買權價值與模擬出來的幾何平均亞式買權價值正相關越高，能降低的變異數越多。

若再考慮 $S_0 - e^{-rT}S_T$ 項，即

$$\widehat{C}^a_{cv} = \widehat{C}^a + \beta_1\left(C^g - \widehat{C}^g\right) + \beta_2\left(S_0 - e^{-rT}S_T\right)$$

則此估計式具不偏性：

$$E\left(\widehat{C}^a_{cv}\right) = E\left(\widehat{C}^a\right) + \beta_1 E\left(C^g - \widehat{C}^g\right) + \beta_2 E\left(S_0 - e^{-rT}S_T\right) = C^a$$

要找到使得變異數最小的 β_1 和 β_2，可以使用 $\widehat{C}^a$ 對 $\widehat{C}^g$ 和 $e^{-rT}S_T$ 作線性迴歸估計。

參數設定：

$$T = 1, \xi_1 = 0.05, S(0) = 50$$
$$m = 16, 64$$
$$\xi_2 = 0.1, 0.3$$
$$K = 45, 50, 55$$

方法：

依給定參數，分別以一般蒙地卡羅法、反向變異法及兩種包含係數的控制變異法，計算亞式買權價值（各模擬 10000 次）。

結果：

```
   S0  K   rf vol period year    mean.C      var.C  mean.C_A     var.C_A
1  50 45 0.05 0.1     16    1 5.9717351  7.7481839 5.9757792  0.04076909
2  50 45 0.05 0.1     64    1 5.9513218  7.8971420 5.9753699  0.04032706
3  50 45 0.05 0.3     16    1 6.8231294 51.2963230 6.8681705  6.91285071
4  50 45 0.05 0.3     64    1 6.9548930 53.5564512 6.9438179  7.01190548
5  50 50 0.05 0.1     16    1 1.8097902  4.4507262 1.8086555  0.67474304
6  50 50 0.05 0.1     64    1 1.8123031  4.4335422 1.8170353  0.69599888
7  50 50 0.05 0.3     16    1 3.9291329 36.0482038 3.9256454 10.10562652
8  50 50 0.05 0.3     64    1 3.9187924 35.0358986 3.9705380 10.03107176
9  50 55 0.05 0.1     16    1 0.1656477  0.4204785 0.1632988  0.19125071
10 50 55 0.05 0.1     64    1 0.1564888  0.3996754 0.1632835  0.19258047
11 50 55 0.05 0.3     16    1 2.0293917 19.5627435 1.9998169  7.71088062
12 50 55 0.05 0.3     64    1 2.0037347 19.7339734 2.0116814  7.77975670
   mean.C_CV      var.C_CV mean.C_CV2     var.C_CV2
1  5.9785206 0.0016550547  5.9784953 0.0014589992
2  5.9763498 0.0014106703  5.9763746 0.0012925530
3  6.9936532 0.0787915213  6.9902652 0.0789482128
4  6.9860225 0.0771493453  6.9850614 0.0764843512
5  1.8234597 0.0009742297  1.8233570 0.0008669158
6  1.8219450 0.0010210815  1.8216838 0.0008886533
7  3.9938350 0.0649368610  3.9929883 0.0610044570
8  3.9753977 0.0580366404  3.9747609 0.0534507806
9  0.1674736 0.0006509195  0.1672112 0.0009903546
10 0.1658175 0.0005818827  0.1656118 0.0009643184
11 2.0518586 0.0557847777  2.0508198 0.0492391543
12 2.0408974 0.0538362235  2.0408861 0.0473735213
```

用反向變異法估出的變異數比一般的蒙地卡羅法低，且用兩種控制變異法所沽出的估計值的變異數降低更多，因此估計值也越準確。

```
realGNP       -5.496801e-01
> library(MASS)
> (fit <- lm.ridge(Employed~Unemployed+Year+ Armed.Forces + Population
 + realGNP,data = as.data.frame(longley.std),
+                  lambda=0.1214735)) #ridge 的參數理論值
                 Unemployed          Year  Armed.Forces    Population
-1.431575e-15 -3.039630e-01  5.668863e-01 -1.721383e-01  1.674119e-01
      realGNP
 5.161751e-01
```