

Management Mathematics

Assignment 2

Due Date: May 10, 2021, 5pm

Please solve the following questions and justify your answer by using Python. **Show all your analysis result in your report.** Upload your “html” file including the **answers** and **Python code** with file name: **MM02_StudentID_Name** to **NTU COOL** by due. The late submission is not allowed.

Note:

- 電腦作業使用 Jupyter Lab/Notebook 完成。
- 文件轉成 HTML 格式，上傳至 NTU COOL 作業區。
- It is highly encouraged to discuss the homework with classmates, but DO NOT COPY programs from others. The copying behavior will result in a reduced score according to the discretion of the teaching assistant.
- NTU COOL 上傳是唯一的繳交方式。不能印出來以紙本繳交或帶隨身碟要求拷貝檔案。
- 我們不只關心你的程式碼，更關心你對問題的理解與詮釋。如果你的作業只有程式碼，沒有任何其他的說明，會被扣大部分的分數。
- 檔案名稱：MM##_StudentID_NAME (eg. MM01_A12345678_李大岩)
- 作業最前面應註明這是哪次作業 (如 MM01)，你的學號與姓名、題目題號清楚標明
- 使用「三明治」答題法: (1)說你要做什麼；(2)程式碼；(3)說你的結果是什麼，以及你的觀察與結論。

Please answer following questions and justify your answer. Show all your works in details.

1. (100%) Linear Regression

Please read the following materials about the linear regression

- Linear Regression by Hand
<http://rebeccaferrell.github.io/CSSS508/Homework/template-HW4-key.html>

(a) (20%) As in the above linkage, simulate and generate the data for independent variable and dependent variable. The sample size $n = 2000$ students.

- Independent variable 1: high school math GPA. Ranges from 2.0-4.0, continuous.
- Independent variable 2: binary indicator variable for whether they took calculus in high school or not (1 = yes, 0 = no). Students who had higher high school math grades were more likely to take calculus in high school.
- Independent variable 3: binary indicator variable for whether they took NTU's precalculus class or not (1 = yes, 0 = no). Students who didn't take calculus in high school or who had lower high school math grades were more likely to take precalculus at NTU first.

- Dependent variable: our outcome is NTU Calculus I GPA grade, which is 0.0 or 0.7-4.0 in 0.1 increments. Let's suppose the following noisy relationship holds for each student i , which we want to recover from our data:

$$\text{NTU_Calculus_I_grade}_i = 0.3 + 0.7 \cdot \text{HS_math_GPA}_i + 0.3 \cdot \text{HS_calculus}_i + 0.1 \cdot \text{NTU_precalculus}_i + \varepsilon_i$$

where $\varepsilon_i \sim \text{Normal}(\text{mean}=0, \text{Standard Deviation}=0.5)$

- (b) (10%) Use the dataset in (a) with the independent and dependent variables given above. Estimate linear regression with “**lm()**” function. Show the results.
- (c) (10%) Setup the matrix X (intercept plus independent variables) and variable y (dependent variable).
- (d) (10%) Compute matrix quantities $(X^T X)^{-1}$ by using “**inv()**” function in NumPy. Let $C = X^T X$. What is the adjoint matrix $\text{adj}(C)$? what is the determinant $\det(C)$? What's the inverse of C ? Hint: <https://www.geeksforgeeks.org/how-to-inverse-a-matrix-using-numpy/>
- (e) (10%) Compute quantities including $\hat{\beta} = (X^T X)^{-1} X^T y$, residuals, residual variance $\hat{\sigma}^2$, standard errors of the covariance matrix $\widehat{\text{Var}}(\hat{\beta})$.
- (f) (10%) Compute the results of the estimated coefficient $\hat{\beta}$. What's the results you observe?
- Make the first column of the matrix have the name “Truth” and contain the true values of the coefficients.
 - Make the second column of the matrix have name “Manual” and the rows be the values of $\hat{\beta}$ you computed.
 - Make the third column of the matrix have name “lm” and the rows be the values of the fitted parameters using **lm()** function from before.
- (g) (10%) Do the similar things in (e), compute the results of the estimated standard errors $\sqrt{\widehat{\text{Var}}(\hat{\beta})}$. What's the results you observe?
- (h) (10%) Do the similar things in (f), compute the results of the estimated residual variance $\hat{\sigma}^2$. What's the results you observe?
- (i) (10%) Based on the results in (f), now, if there is a new student, who has $\text{HS_math_GPA} = 3.123456$, $\text{HS_calculus} = 1$, and $\text{NTU_precalculus} = 0$. What's his/her prediction of $\text{NTU_Calculus_I_grade}$ according to your linear regression Truth, Manual, and lm, respectively?

Note

1. Show all your work in detail. Innovative idea is encouraged.
2. If your answer refers to any external source, please “must” give an academic citation. Any “plagiarism” is not allowed.