Management Mathematics
Assignment 3
Due Date: June 2, 2022, 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: MM03_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. MM03_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. **(70%)** Cosine Similarity

   Please read the following materials about the Cosine Similarity

   - Cosine Similarity
     https://danushka.net/lect/dm/Numpy-basics.html

   (a) **(10%)** As in the above linkage, we generate six sentences are:

   - A = "I love Management Mathematics"
   - B = "I hate Management Mathematics"
   - C = "I love Operations Research"
   - D = "I hate Operations Research"
   - E = "I love Manufacturing Data Science"
   - F = "I hate Manufacturing Data Science"

   (b) **(10%)** Based on the data in (a), get the list of words in each sentence by calling the split() in strings, and convert the sentence into lowercase by using lower(). Show the results.

   (c) **(10%)** Use the set class to identify the dimensionality of the feature space and print all

vocabulary.

(d) (**10%**) To build the one-to-one mapping between features and IDs in the feature space. We use the list() to convert the vocabulary into unique features.

(e) (**10%**) Declare six 1D-arrays for the six sentences and build each word vector by increasing the corresponding feature value (Hint: that is to build vA, vB, vC, vD, vE, and vF).

(f) (**10%**) Compute the cosine similarity between two vectors, and show a 6x6 matrix including these similarity values (Hint: the diagonal elements should be 1).

(g) (**10%**) Based on the similarity matrix, what's the results you draw?

2. (**30%**) Information Retrieval

Please follow the course lecture notes (Chapter 4) to finish the information retrieval work.

(a) (**10%**) Build up the following matrix for preparing information retrieval, and convert all column vectors into unit vectors as matrix **Q**.

Table 1	Frequency of Key Words							
	Modules							
Key words	M1	M2	M3	M4	M5	M6	M7	M8
determinants	0	6	3	0	1	0	1	1
eigenvalues	0	0	0	0	0	5	3	2
linear	5	4	4	5	4	0	3	3
matrices	6	5	3	3	4	4	3	2
numerical	0	0	0	0	3	0	4	3
orthogonality	0	0	0	0	4	6	0	2
spaces	0	0	5	2	3	3	0	1
systems	5	3	3	2	4	2	1	1
transformations	0	0	0	5	1	3	1	0
vector	0	4	4	3	4	1	0	3

(b) (**10%**) To do a search for the key words *determinants*, *matrices*, and *systems*. Form a unit search vector **x** whose entries are all 0 except for the three rows corresponding to the search rows.

(c) (**10%**) Calculate the cosine similarity $\mathbf{y} = \mathbf{Q}^T\mathbf{x}$, and rank the modules.

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.