FINAL REPORT

Course: Applied Linear Algebra for IT

Course Code: 501032

I. REGULATIONS

This report is a replacement for the final written exam in the second semester of the academic year 2024-2025. Students need to adhere to the following regulations:

• Each report is conducted by a group of **one or two students**.

• For each question, students are required to present **detailed calculation steps**, and the final results.

• The report must be submitted in **PDF format**, and the content must be written based on the report/essay format of the Faculty of Information Technology. **In case students do not follow the Faculty's format, they will receive 0 points for the Report.**

• Each group of students must complete the **self-evaluation form** using the following template:

- https://docs.google.com/spreadsheets/d/1WkFQLo6uvYalRSqT9dbsryC9QFjScKIV/edit?usp=sharing&ouid=104602653034506388459&rtpof=true&sd=true
- Download the form as an Excel file (.xlsx extension)

• The report must include the following contents:

a. Chapter 1: Solutions (9.0 points)

Write **detailed calculation steps** and the final results of each question (in the section III). In each calculation step, insert the related Background knowledge (theorems, laws, rules, tests…) in case you apply it, for ex.,:

- By Gaussian Elimination algorithm, ...

• Unusual cases (such as identical patterns, signs of plagiarism, etc.) will be scheduled for online discussions with the instructor.

II. Submission guideline

- Filenames of the report and self-evaluation files must be the **Student IDs**, for ex.,
  - A group of only one student with student ID 521H1495 will submit a self-evaluation file named **521H1495.xlsx** and a report file named **521H1495.pdf**
  - A group of two students with student IDs 521H1234 and 522H4321 will submit a self-evaluation file named **521H1234_522H4321.xlsx** and a report file named **521H1234_522H4321.pdf**
- Each student in a group submits **two files** to the "**FinalReport**" assignment on Elearning website of the Theory class.

III. Questions

Question 1 (9.0 points). Investigate in-detail the singular value decomposition (SVD) of a rectangular matrix:

- Definition/Formulas of singular values
- Definition/Formulas of SVD
- Steps for finding a SVD
- Five applications of SVD
- Take two examples of SVD finding (examples from Internet/books **are not allowed**) :
    - One with a matrix of size 2 rows × 3 columns and the rank of the matrix is 2
    - One with a matrix of size 2 rows × 3 columns and the rank of the matrix is 1
- Definition/Formulas of compact (reduced) SVD
- Find the compact SVDs of the two input matrices in the two previous examples
- Definition/Formulas of truncated SVD
- Find the truncated SVD of the input matrix of rank 2 in the previous examples, and we keep only **one nonzero singular value** in the truncated SVD of the input matrix
- Apply truncated SVD to the following recommendation system:

	Transformers	**The Avengers**	**Titanic**	**Casablanca**
Liam	5	5	1	
Noah	4	5	1	1
Oliver	4	4	1	2
James	5	4	1	1
Olivia	1	2	5	5
Emma		1	5	5
Mia		1	5	5
Luna	5	4	1	2

We have a table of data, like a set of ratings where rows are people, and columns are movies. The numbers in the table show how much each person likes each movie. Missing values mean that the people have not yet rated the corresponding movies. There are **three missing values** in the table (red cells). Apply truncated SVD to the rating data above to predict proper values for the missing data by two following approaches:

Step order	**Step content**	**Approach 1** **Non-centering data**	**Approach 2** **Centering data**
1	Create the input matrix from the rating data	✓	✓

2	Fill each missing value by the average value of row mean and column mean	✓	✓
3	**Centering the data by row**: subtract the mean of each row from the corresponding values in each row		✓
4	Find the truncated SVD of the modified matrix above: keep only **one nonzero singular value**	✓	✓
5	Reconstruct an approximation of the input matrix by the truncated SVD	✓	✓
6	Original means recovering: add the original mean of each row (calculated in Step 3) from the corresponding values in each row		✓

After finishing these steps, we obtain the final results which are recommendations for people including predicted values for missing data.

Do two above approaches and give conclusions and explanations for the obtained results:

- The obtained results from the final approximation matrix still properly present the preferences of each person or not? Why?
- Predicted values for missing data are good and relevant to people's preferences? Why?

(See Appendix A for a short guideline of solving the SVD-based recommendation system)

IV. Appendix A

This is a short guideline of solving the SVD-based recommendation system. You must present **detailed calculation steps** in this final report.

In this guideline, we only handle the "Centering data" approach (the "non-Centering data" one is similar):

Example rating data:

	Movie 1	**Movie 2**	**Movie 3**
Person 1	5	5	
Person 2	4	5	1
Person 3	1	4	4
Person 4	5	4	1

Step order	**Step content**	**Approach 2** **Centering data**

1	Create the input matrix from the rating data	[[5. 5. **?**] [4. 5. 1.] [1. 4. 4.] [5. 4. 1.]]
2	Fill each missing value by the average value of row mean and column mean	For the missing value (? symbol): Row mean = (5 + 5) / 2 = 5 Column mean = (1 + 4 + 1) / 3 = 2 Thus, “?” → (row mean + column mean) / 2 = (5 + 2) / 2 = 3.5 [[5. 5. **3.5**] [4. 5. 1.] [1. 4. 4.] [5. 4. 1.]]
3	**Centering the data by row**: subtract the mean of each row from the corresponding values in each row	Mean value of each row = [[4.5] [3.33333333] [3.] [3.33333333]] Mean subtraction = [[5 - 4.5 5 - 4.5 **3.5** - 4.5] [4-3.33333333 5-3.33333333 1-3.33333333] [1 - 3 4 - 3 4 - 3] [5-3.33333333 4-3.33333333 1-3.33333333]] = [[0.5 0.5 -1.] [0.66666667 1.66666667 -2.33333333] [-2. 1. 1.] [1.66666667 0.66666667 -2.33333333]]
4	Find the truncated SVD of the modified matrix above: keep only **one nonzero singular value**	$\hat{U}_{4\times1}$ = [[-0.27077534] [-0.60341643] [0.35595345] [-0.66020184]] $\hat{\Sigma}_{1\times1}$ = [[4.45030224]] $\hat{V}^T_{1\times3}$ = [[-0.52803348 -0.27532126 0.80335474]]
5	Reconstruct an approximation of the input matrix by the truncated SVD	Approximation matrix = [[0.6362973 0.33177096 -0.96806827] [1.41797345 0.73934372 -2.15731717] [-0.83645808 -0.43613653 1.27259461] [1.55141397 0.80892077 -2.36033474]]
6	Original means recovering: add the original mean of each row (calculated in Step 3) from the corresponding values in each row	Final approximation matrix = [[5.1362973 4.83177096 3.53193173] [4.75130679 4.07267705 1.17601616] [2.16354192 2.56386347 4.27259461] [4.8847473 4.1422541 0.9729986]]

THE END