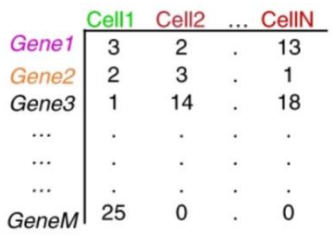
BCS410. Practical Class 7 & Homework 6 [Due: June 2]

Background: A single-cell RNA-seq dataset: A single-cell RNA-seq dataset typically captures gene expression profiles at the resolution of individual cells. The data is often structured as a matrix with genes as rows and cells as columns, where each entry represents the number of mRNA transcripts (counts) observed for a particular gene in a particular cell.



	Cell1	Cell2	...	CellN
Gene1	3	2	.	13
Gene2	2	3	.	1
Gene3	1	14	.	18
...	.	.	.	.
...	.	.	.	.
...	.	.	.	.
GeneM	25	0	.	0

As described above, single-cell RNA sequencing (scRNA-seq) allows the measurement of gene expression at the level of individual cells. One of the main tasks in analyzing such data is to cluster cells based on their gene expression profiles, enabling the identification of distinct cell types or states.

Problem 1: Import the given dataset 'Preprocessed_Single_Cell_RNA_Seq_data.csv'. Then, perform PCA to reduce the data to two dimensions. Based on your visual inspection, estimate how many distinct cell types are present in the dataset. Please refer to "PCA_skeleton_code.ipynb".

Problem 2: Perform PCA to reduce the data to three dimensions. Based on your visual inspection, estimate how many distinct cell types are present in the dataset. Please refer to "PCA_skeleton_code.ipynb".

Problem 3: Perform t-SNE to reduce the data to two dimensions. Based on your visual inspection, estimate how many distinct cell types are present in the dataset. Please refer to "t_SNE_skeleton_code.ipynb".

Problem 4: Perform UMAP to reduce the data to two dimensions. Based on your visual inspection, estimate how many distinct cell types are present in the dataset. Please refer to "UMAP_skeleton_code.ipynb".

Bonus problem 1 (5 bonus points): Compare the computational time cost among PCA, t-SNE, and UMAP. Hint: You can use the library named "time".