

Fall 2025: CAP5610 – HW 2

In HW1, you will learn dimensionality reduction and data visualization tools: PCA and t-SNE, and Data visualization tool: Violin plot.

Data:

- Gene Expression: lncRNA_5_Cancers.csv (available under Module 2)
- Cancer Types: KIRC, LUAD, LUSC, PRAD, THCA

Task 1: [25 points] Visualize the lncRNA expression data of five different cancer types using PCA. Reduce the data dimension from 12,309 to two (PC1 and PC2) dimensions and plot the data into reduced dimensions. Must plot all the data of five cancer types.

Task 2: [25 points] Draw two violon plots – one with the values of PC1 and the other with PC2.

Task 3: [25 points] Repeat task 1 using t-SNE library. Plot the data in reduced dimension using two t-SNE components (t-SNE 1 and t-SNE 2).

Task 4: [25 points] Draw two violon plots – one with the values of t-SNE 1 and the other with t-SNE 2.

You must submit the following items in CANVAS:

- Report (MS word or PDF)
 - Describe the algorithms/approaches/tools used: (a) What it is or What it does, (b) How it does, and (c) Application.
 - Describe results: (a) Put Figure/Table number and Title: On top of the table, and bottom of the figure. (b) Describe the figure and table. (c) Your observation about the figure and table. (d) Conclusion.
- Source code (*.py or Jupyter notebook)
 - Must be well organized (comments, indentation, ...)
- File name: HW2_lastName

You must submit the files **SEPERATELY**. DO NOT compress into a ZIP file. If you fail to provide all required information or files, you may be given zero score without grading.

Deadline:

The deadline is **11:59pm Wednesday, September 17, 2025**. Late assignments will not be accepted.