IITM-CS5691 : Pattern Recognition and Machine Learning — Release Date: October 9, 2023
Assignment II — Due Date : October 23, 2023, 23:59

Roll No: CS23E001 — Name: Shuvrajeet Das
Collaborators (if any):
References/sources (if any):

- Use LaTeX to write-up your solutions (in the solution blocks of the source LaTeX file of this assignment), submit the resulting rollno.asst2.answers.pdf file at Crowdmark by the due date, and propery drag that pdf's answer pages to the corresponding question in Crowdmark (do this propery, otherwise we won't be able to grade!). (Note: **No late submissions** will be allowed, other than one-day late submission with 10% penalty or four-day late submission with 30% penalty.)
- Please upload to moodle a rollno.zip file containing three files: rollno.asst2.answers.pdf file mentioned above, and two code files for the programming question (rollno.ipynb file and rollno.py file). Do not forget to upload to Crowdmark your results/answers (including Jupyter notebook **with output**) for the programming question.
- Collaboration is encouraged, but all write-ups must be done individually and independently, and mention your collaborator(s) if any. Same rules apply for codes written for any programming assignments (i.e., write your own code; we will run plagiarism checks on codes).
- If you have referred a book or any other online material or LLMs (Large Language Models like ChatGPT) for obtaining a solution, please cite the source. Again don't copy the source *as is* - you may use the source to understand the solution, but write-up the solution in your own words (this also means that you cannot copy-paste the solution from LLMs!). Please be advised that *the lesser your reliance on online materials or LLMs* for answering the questions, *the more your understanding* of the concepts will be and *the more prepared you will be for the course exams*.
- Points will be awarded based on how clear, concise and rigorous your solutions are, and how correct your answer is. The weightage of this assignment is 12% towards the overall course grade.

1. (8 points) [PRINCIPAL COMPONENT ANALYSIS - NUMERICAL] Consider the following dataset D of 8 datapoints:

   You need to reduce the data into a single-dimension representation. You are given the first principal component: PC1 $= (-0.694, -0.720)$.

   (a) (2 points) What is the xy coordinate for the datapoint reconstructed (approximated) from data #2 (x=20.82, y=24.03) using the first principal component of D? What is the reconstruction error of this PC1-based approximation of data #2?

data #	x	y
1	5.51	5.35
2	20.82	24.03
3	-0.77	-0.57
4	19.30	19.38
5	14.24	12.77
6	9.74	9.68
7	11.59	12.06
8	-6.08	-5.22

Solution:

$$\text{Projected value} = \text{data point} \cdot \text{principal component}$$

For data point 2:

$$\text{Projected value} = (20.82, 24.03) \cdot (-0.694, -0.720) = -0.694 \times 20.82 - 0.720 \times 24.03 = -31.3708$$

To reconstruct the original data point from its projection onto the first principal component, you can use the following formula:

$$\begin{aligned}
\text{Reconstructed data point} &= \text{Projected value} \times \text{Principal component} \\
\text{Reconstructed data point} &= -31.3708 \times (-0.694, -0.720) \\
\text{Reconstructed data point} &= (21.77849, 22.606656)
\end{aligned}$$

The reconstruction error is:

$$\begin{aligned}
\text{Reconstruction error} &= \sqrt{(20.82 - 21.78)^2 + (24.03 - 22.61)^2} \\
\text{Reconstruction error} &= \sqrt{(-0.96)^2 + (1.42)^2} \\
\text{Reconstruction error} &= \sqrt{2.1476 + 2.0164} \\
\text{Reconstruction error} &= \sqrt{4.164} \\
\text{Reconstruction error} &\approx 2.04
\end{aligned}$$

The reconstruction error of this PC1-based approximation of data #2 is approximately 2.04 units.