Week-8 PCA

DS3010: Introduction to Machine Learning Lab

Timing: 02:00 PM to 04:30 PM Max Marks: 5

Instructions

1. Submit one .ipynb file containing all answers. The name should be [student_name] _[rollno]_[lab].ipynb
2. Write the questions in separate text blocks before the answers.
3. Outputs for all sub-questions should be provided, and the code should be executable.
4. Write justifications for your choices where needed.
5. Ensure that all plots include clear labels and legends for better interpretation.
6. Use of generative AI tools (such as ChatGPT, Gemini, etc.) is strictly prohibited. Any submission found to contain AI-generated or plagiarized content will receive a score of zero and may face disciplinary action.

Load the given *heart.csv* file.

1. **Data Preparation** (0.5)
    1. Do the necessary pre-processing task, if needed.
    2. Split the data using the train test split (8:2).
2. **Classification using custom PCA** (2.5)
    - Centralize the data.
    - Calculate the covariance matrix.
    - Calculate the eigenvalues and eigenvectors.
    - Sort the eigenvectors in descending order based on eigenvalues.
    - Project the data into lower dimensions and calculate the reconstruction loss.
    - Plot the graph for reconstruction loss vs number of components.
    - Select the optimal number of components from the plot and perform PCA to get the reduced feature vectors.
3. Perform classification without PCA and with PCA with any model of your choice. Print the classification report. (1.5)
4. Analyze the results and write your observation. (0.5)