

Indian Institute of Technology Jodhpur
CSL2010: Introduction to Machine Learning
Lab 8&9, Due Date: Oct 17 , 2025, Max Marks: 70+30 for Viva

1. Consider the datasets: <https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data>. Each point of this dataset is a 4-dimensional vector ($d = 4$) given in the first four columns of the datafile. Assume that you want to reduce the dimensionality to k ($k = 2$). This dataset contains 3 clusters. Ground-truth cluster IDs are given as the fifth column of the data file.
 - (a) Read the dataset and mean center each of the feature. Let matrix $X = [x_1 \ \dots \ x_n] \in \mathbb{R}^{d \times n}$ represents your dataset where each column represents a data-point.
[Compulsory to implement in the lab 8].
 - (b) Find the data covariance matrix $C = \frac{1}{n}XX^T$ and its eigenvalue decomposition as $C = V\Sigma V^T$ where $V = [v_1 \ \dots \ v_d] \in \mathbb{R}^{d \times d}$ contains the eigenvectors and $\Sigma = \text{diag}([\lambda_1 \ \dots \ \lambda_d])$ contains the eigenvalues where $\lambda_1 \geq \dots \geq \lambda_d$.
[Compulsory to implement in the lab 8].
 - (c) Create the projection matrix $W \in \mathbb{R}^{k \times d}$ of the principal component analysis algorithm that project d dimensional data-point to a point in k dimensional space. Find the lower dimensional representation $y_i = Wx_i$ of each of the original datapoint.
 - (d) Create the reconstruction matrix $U \in \mathbb{R}^{d \times k}$ that reconstruct the original data point x_i from its lower dimensional representation y_i . Let $\hat{x}_i = Uy_i$ be the reconstructed point.
 - (e) Find the reconstruction error $\frac{1}{n} \sum_{i=1}^n \|x_i - \hat{x}_i\|_2^2$
 - (f) Perform clustering (in 3 clusters) before and after dimensionality reduction using the Spectral Clustering algorithm and then find the percentage of points for which the estimated cluster label is correct. Report the accuracy of the Spectral Clustering algorithm before and after the dimensionality reduction. Report the reconstruction error for $k = 1, 2, 3$.
[Compulsory to implement in the lab 9].
 - (g) Plot the obtained cluster.
[Compulsory to implement in the lab 9].