

**Lab - 7**  
**CSL2010: Introduction To Machine Learning**  
**AY 2022-23**

**Principal Component Analysis**

**(Due: 28 Sep 2022, 11:59 PM)**

**General Instructions**

1. You need to upload a zip **<Lab7\_Your\_Roll\_No>.zip**, which contains one file for the task in **<Lab7\_Your\_Roll\_No>.py** format and the report for the entire assignment in **<Lab7\_Your\_Roll\_No>.pdf** format.
2. Provide your colab file link in the report. **Make sure that your file is accessible.**
3. Submit a single report mentioning your observations for all the tasks.
4. Report/Cite any resources you have used while attempting the assignment.
5. Attempt (1), (2) and (3) during the lab.

**[Principal Component Analysis] :- [35 Marks]**

**Dataset**

1. From the given link, download "iris.data", "iris.names" and convert them into a readable format (Ex: txt, csv, etc.), perform some meaningful Exploratory Data Analysis, and perform the pre-processing after splitting into train and test sets. **[5 Marks]**
2. Choose any one classification method taught in the earlier classes and calculate its accuracy on the test data. **[5 Marks]**
3. Apply the PCA algorithm on the given features. **[5 Marks]**
4. Use the top few principal components to project the data to a certain chosen dimensions, and re-evaluate the above classification method (from part 2). **[10 Marks]**
5. Examine the accuracy for different values of projection dimensions, and compare them. **[10 Marks]**

**Practice Question:** Implement Principal Component Analysis from scratch, repeat the above experiments in (4) and (5), and compare the two results. You may utilize appropriate functions from the NumPy library in your implementation for computing eigenvectors and eigenvalues.