AIM 511 Machine Learning Project Requirements

---

1 Project

1.1 Overview

- This project accounts for 15% of the final grade.
- Group Size: max. 3 members
- Total marks: 25

1.2 Problem Statements

- Develop a machine learning model to predict if a client will subscribe (yes/no) to a term deposit. Bank has multiple banking products that it sells to customers, such as savings account, credit cards, investments etc. It wants to determine which customer will purchase its credit cards. For the same it has various kind of information regarding the demographic details of the customer, their banking behavior etc. Dataset
- Develop a machine learning model to predict traffic volume in Bangalore using the given dataset . Begin by performing clustering on the dataset to identify patterns or segments that might enhance model accuracy. Evaluate prediction model, comparing performance before and after incorporating clustering results to determine if clustering improves prediction capabilities. Dataset
- Develop a machine learning model to predict the number of bikes rented at different times, leveraging features such as weather conditions, time of day, and holiday status. Additionally, use clustering techniques to identify patterns in demand, grouping similar periods together to uncover insights into peak times, low-demand periods, and the impact of different conditions on bike usage. These insights can help optimize the bike-sharing system's operations and improve resource allocation. You will use the `hour.csv` dataset for detailed hourly predictions and the `day.csv` dataset for analyzing broader daily trends. The insights gained from this analysis will help improve the operational efficiency of the bike-sharing system and enhance user satisfaction. Dataset
- Develop a machine learning model to predict water temperature. The CalCOFI data set represents the longest (1949-present) and most complete (more than 50,000 sampling stations) time series of oceanographic and larval fish data in the world. It includes abundance data on the larvae of over 250 species of fish; larval length frequency data and egg abundance data on key commercial species; and oceanographic and plankton data. The physical, chemical, and biological data collected at regular time and space intervals quickly

became valuable for documenting climatic cycles in the California Current and a range of biological responses to them. CalCOFI research drew world attention to the biological response to the dramatic Pacific-warming event in 1957–58 and introduced the term "El Niño" into the scientific literature. Use `bottle.csv`.Dataset

- This project involves working with the `moonDataset.csv`, which contains 200 data points across four columns. The features X1 and X2 are derived from a 2D 'moons' shape, a common structure used for testing non-linear classification algorithms, while X3 introduces a vertical displacement, creating a 3D effect and adding complexity to the dataset. The label column provides a binary class for each data point, indicating whether it belongs to class 0 or 1. Your task is to perform non-linear classification using Gaussian Mixture Models (GMMs). Begin by exploring and visualizing the dataset to understand the 3D structure and the distribution of data points. Then, implement GMMs to classify the data points into their respective classes based on the provided features, analyzing the model's performance and effectiveness in handling non-linear boundaries.Dataset
- Given a dataset containing information about candidate attributes, assessment scores, and hiring decisions, develop a predictive model that accurately determines whether a candidate will be hired or not. The model should be able to identify key factors influencing hiring decisions and assist in optimizing the recruitment process for improved efficiency and equity. Dataset

The choice of deciding to do a project or paper must be updated in the sheet linked on LMS by **30th August 2024, 11:59PM**. Failure to do so will incur a penalty during the evaluation. No more than 10 teams will be allowed to choose the same problem statement.

1.3 Submission requirements

Each team that chooses the project track will have to submit the notebook with the code. The notebook should contain all work that the team wishes to consider.

1.4 Evaluation Criteria

- Code: 10 marks
  - ! EDA and inferences: 4 marks
  - ! Preprocessing: 3 marks
  - ! Models and results: 3 marks
- Viva: 15 marks

2 Research Paper

You can alternatively choose to write a research paper on at least **two** Machine learning concepts/topics discussed in the class so far. They have to be original and should not be plagiarized or AI generated.

The number of pages should be atleast SIX pages (without references) in the double-column single-spaced IEEE conference format in Latex: `https://www.ieee.org/conferences/publishing/templates.html`.

2.1 Evaluation Criteria

The paper will be evaluated based on

- Originality
- Clarity
- Analogies
- Correctness

3 Important dates

– **30th August 2024, 11:59 PM**: Finalize paper/project and update on the sheet.

– **13th October 2024, 11:59 PM**: Final submission of report/paper.