Department of Computer Science

University *of* Kerala

M Sc Computer Science (AI) 24 - 26 Batch

CAI-CC-524-Pattern Recognition Laboratory

Cycle 1- List of Experiments

1. Identify the most suitable patterns in the given a set of data.

#	Sequence					
1	2	4	6	8	10	
2	3	6	12	24	48	
3	1	1	2	3	5	8
4	10	20	30	55	70	
5	100	50	25	12.5	6.25	

   Write a python program to find the exact pattern in the given series of data.

2. Create a textdata.csv file containing a minimum 10 sentences with a minimum length of 15 words and find the following

   - Most common words
   - Most common bigrams and trigrams
   - Palindrome word
   - Anagram pairs and rhyming words.

   Repeat the experiments with any two articles of two authors and find any patterns exists in it.

3. Write a program to detect edges in an image using the Canny edge detection algorithm, where the program should read an image in grayscale, apply appropriate threshold values, and display both the original and edge-detected images.

4. Create images with different non overlapped filled shapes. Implement K-means clustering for colour segmentation using those images. Display the shapes separately.

5. Consider UCI Wine Quality Dataset, which contains physicochemical properties of wines along with their quality scores (ranging from 0 to 10). Your task is to analyze the statistical properties (mean, variance, skewness, and entropy) of the given features and use them to find out

   - wine quality.
   - Which statistical measure is the most important predictor of wine quality?
   - Does high entropy indicate harder classification of wine quality?
   - What happens if we drop features with high variance?

6. Consider the SMS Spam Collection dataset, implement a python program to analyze the syntactical patterns (word frequency, sentence length, POS tagging distribution) and investigate their role in classifying messages as spam or ham (non-spam).

7. Collect Combined Cycle Power Plant (CCPP) dataset, which contains 9568 observations from UCI repository. Perform exploratory data analysis (EDA) to understand the data distribution, detect missing values, and visualize relationships between features and the target variable. Preprocess the data. Split the dataset into training (80%) and testing (20%) sets. Implement at least two regression models, such as Linear Regression, Decision Tree Regression, Random Forest Regression, or any other advanced technique of your choice. Evaluate the models using appropriate regression performance metrics. Compare the performance of the models and justify which model is more suitable for predicting power output.
8. You are provided with the Breast Cancer Wisconsin (Original) Dataset from the UCI Machine Learning Repository. This dataset contains 699 instances with 10 attributes describing the cellular characteristics of breast biopsies. Perform Exploratory Data Analysis (EDA). Preprocess the dataset. Convert the target labels from $2 \rightarrow 0$ (Benign) and $4 \rightarrow 1$ (Malignant). Normalize or standardize features if necessary. Split the dataset into training (80%) and testing (20%) sets. Implement at least two classification models such as Logistic Regression, Decision Tree Classifier, Random Forest Classifier, Support Vector Machine (SVM). Evaluate the models using appropriate performance metrics. Compare the performance of the models and justify which model is more suitable for breast cancer classification.
9. Download a real-world dataset containing at least 1000 samples and more than 10 features. Perform exploratory data analysis (EDA) to understand its structure. Apply Principal Component Analysis (PCA) to reduce dimensionality and analyze variance explained by principal components. Visualize the dataset using different sets of principal components and interpret the results.
10. Find out a real-world dataset containing at least 1000 samples and more than 10 features with class labels. Perform exploratory data analysis (EDA) to understand the dataset's structure. Apply Linear Discriminant Analysis (LDA) for dimensionality reduction, ensuring class separability. Visualize the transformed dataset using different linear discriminants and analyze the effectiveness of LDA in improving class separation
11. Design and implement perceptrons, write you own code for learning (weight updating) task, for the AND, OR, NAND, and NOR logic gates. Conduct experiments to evaluate the impact of bias on the perceptron's ability to learn and classify these logic gates accurately
12. A shopping mall wants to segment its customers based on their spending patterns. You are given a dataset containing customer details, including Annual Income and Spending Score. Your task is to apply K-Means Clustering to group customers into meaningful segments and analyze the results. Use Mall Customers Dataset from Kaggle.

Basic Level Questions	1, 2, 3,4,11
Advanced Level Questions	5,6,7,8,9,10,12