MACHINE LEARNING

Course Code: ISL66 **Credits: 0:0:1**
Prerequisites: Scripting Languages **Contact Hours: 14P**
Course Coordinator: Dr.Mydhili K Nair

Laboratory Experiments for Batch 2022 - 2023

Part A

Implement the following programs using Python without using Built In functions.

1. **Model Measurement Analysis**: Using ***any dataset*** and ***any classifier***, calculate TP, TN, FP and FN from sklearn library functions. Also calculate different metrics (Accuracy, Precision, Recall(Sensitivity), F1-Score, MCC, Specificity, Negative Predictive Value) by defining our own functions. Compare your values with scikit-learn's library functions. Get the result of Confusion Matrix using sklearn. Using sklearn, plot the ROC & AUC Curves for your test data and random probabilities. Using sklearn, calculate the AUC of your test data and of random probabilities. Interpret the results. Write the inference/analysis of each output. Expectation is Supervised Learning Algorithms. Therefore, use any classifier.
2. **Supervised Learning Algorithms - KNN:** Build a KNN model for predicting if a person will have diabetes or not with a high accuracy score. Perform some appropriate Pre-Processing steps on the given dataset for better results. Implement the KNN algorithm on your own. Try other possible processes that can be done to dataset and tuning the model to increase accuracy such as Increase K value, Normalization and Different Distance Metrics. Perform Feature Ablation Study. Additional Tries: Weight the features before doing KNN prediction.
3. **Un-Supervised Learning Algorithms - K-Means Clustering:** Build a K-Means Model for the given dataset. Build a K-Means Model for the given Dataset. Use the Elbow or Silhouette method to find the best possible K-value. Use the cost function as WCSS(Within Cluster Sum of Squares). Curate a new data point and find which cluster it goes to. Write the inference.

Part B

Implement the following programs using Python.

1. **Supervised Learning Algorithms - Simple Linear Regression (Univariant):** Consider ***any dataset*** from UCI repository. Create Simple Linear Regression models using the training data set. Predict the scores on the test data and find the error in prediction (E.g. RMSE, MAE, LSE). Include appropriate code snippets to visualize the model. Use Sub-Plots Interpret the result. Write the Inference.

2. **Supervised Learning Algorithms - Multiple Linear Regression (Multivariate):** Consider ***any dataset*** from UCI repository. Create Multiple Linear Regression models using the training data set. Predict the scores on the test data and find the error in prediction (E.g. RMSE, MAE, LSE). Include appropriate code snippets to visualize the model. Use Sub-Plots Interpret the result. Write the Inference.
3. **Probabilistic Supervised Learning - Naive Bayes(Binomial):** Create a dataset from the sample given to you(e.g. "Titanic, Play Tennis Probability", "Shopper Buying Probability" etc.). Perform the necessary pre-processing steps such as encoding. Train the model using Naive Bayes Classifier for Binomial predictions. Give new test data and predict the classification output. Handcode the classification probability and compare with the model output. Analyze and write the inference.
4. **Probabilistic Supervised Learning - Naive Bayes(Multinomial):** Create a dataset from the sample given to you(e.g. "Iris", "Advertising" etc.). Perform the necessary pre-processing steps such as encoding. Train the model using Naive Bayes Classifier for Multinomial predictions. Give new test data and predict the classification output. Handcode the classification probability and compare with the model output. Print the confusion matrix, accuracy, precision, recall, sensitivity, specificity, MCC etc. Analyze and write the inference.
5. **Artificial Neural Networks - Multi Layer Perceptron:** Write a program to use a Multi Layer Perceptron from sklearn for classifying an outcome for any dataset. The program should contain functions to read the dataset from the file, pre-process it for encoding categorical attributes and split the data into train and test. Train the model. Predict the output sample test data. Perform Mini-Batch Training and plot the accuracy score for the train and test data. Print the confusion matrix, accuracy, precision, recall, sensitivity, specificity, MCC etc. Write the Inference.
6. **Supervised Learning Algorithms - Decision Trees:** Implement decision trees considering a data set of your choice. Create an ID3 Decision Tree. Implement the ID3 algorithm on your own. Use a pre-defined sklearn library for both Entropy-Information Gain as well as Gini Index calculation. Compare the results you got between your hand-coded algorithm and sklearn. Draw the decision tree. Write the interference on the choice of the root node, the internal nodes and the leaf nodes. Also calculate the Decision Tree model measurement metrics such as confusion matrix, accuracy, precision, recall, sensitivity, specificity, MCC etc.Write the inference.
7. **Supervised Learning Algorithms - Logistic Regression (Univariant):** Implement logistic regression and test it using any dataset which will be given to you. Give new test data and predict the classification output. Handcode the classification probability and compare with the model output. Print the confusion matrix, accuracy, precision, recall, sensitivity, specificity, MCC etc. Analyze and write the inference.

8.	**Un-Supervised Learning Algorithms - Hierarchical Clustering:** Using any dataset implement Hierarchical Clustering (AGNES and DIANA). Plot the Dendrogram for Hierarchical Clustering and analyze your result. Plot the clustering output for the same dataset using these two hierarchical techniques. Compare the results. Write the inference.
Reference:	
1.	1. Stephen Marsland, "Machine Learning - An Algorithmic Perspective",Second Edition, CRC Press - Taylor and Francis Group, 2015
2.	2. Ethem Alpaydin, "Introduction to Machine Learning", Second Edition, MIT Press, Prentice Hall of India (PHI) Learning Pvt. Ltd. 2010
Course Outcomes (COs):	
At the end of the course, student will be able to -	
1.	Design the experiment for the given problem using various Machine Learning Algorithms **(PO-1,2,3,5,6,9,10,12) (PSO-3)**
2.	Develop the solution for the given real world problem**(PO-1,2,3,4,5,6,8,9,10,12) (PSO-3)**
3.	Analyze the results and produce substantial written documentation **(PO-1,2,4,9,10) (PSO-3)**

Conduction of Practical Examination: - (50 Marks)

- All laboratory experiments are to be included for practical examination.
- Marks Distribution:
  - ➢ Procedure Writing (20 Marks)
  - ➢ Implementation and Testing (20 Marks)
  - ➢ Viva (10 Marks)