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

**Laboratory Experiments:**

| **Implement the following programs using Python** | |
| --- | --- |
| 1. **Model Measurement Analysis**: Assume a black box model for a Binary Classifier that categorizes any dataset with a categorical attribute and having at least 10 records. **E.g**. CoronaVirus patients who were tested **(or)** Student Assignments subjected to plagiarism check etc. The model is fed the values of TP, TN, FP, FN ensuring a total sample size of 100. We can alter the threshold for labeling to maximize the classifier performance. Each time the threshold is increased in steps of 0.1. At each step of varying thresholds calculate the values of Precision, Recall, F1 Score as well as the TPR and FPR. Plot the ROC Curve. Analyze, Interpret. 2. **Artificial Neural Networks - Single Layer Perceptron:** Implement a Single Layer Perceptron using minimal inbuilt functions. Create a dataset containing at least 100 records. Each record should have at least 4 floating point features and a binary label (0 - negative or 1 - positive). Split the dataset into test and train data, initialize the weights, learning rate, epochs and define the activation function. Train the model (Learn the weights of the perceptron on the training data). Print the learned weights and the hyperparameters (epoch and learning rate). Predict the outputs on train and test data. Print the confusion matrix, accuracy, precision, recall on train and test data 3. **Artificial Neural Networks - Multi Layer Perceptron:** Build an Artificial Neural Network by implementing the Back Propagation Algorithm. Test the same using appropriate data sets. Compare the actual and predicted output. Analyze and write the inference. 4. **Supervised Learning Algorithms - Decision Trees:** Implement decision trees considering a data set of your choice. 5. Create a ID3 Decision Tree 6. Create a CART Decision Tree 7. Compare and Contrast the two 8. **Supervised Learning Algorithms - Linear Regression:** Consider a dataset from UCI repository. Create a **Simple** and **Multiple** Linear Regression model using the training data set. Predict the scores on the test data and output RMSE and R Squared Score. Include appropriate code snippets to visualize the model. Interpret the result. 9. **Supervised Learning Algorithms - Logistic Regression:** Implement logistic regression and test it using any dataset of your choice from UCI repository. The output should include Confusion Matrix, Accuracy, Error rate, Precision, Recall and F-Measure. 10. **Supervised Learning Algorithms - KNN:** Implement k-Nearest Neighbor (KNN) by writing the algorithm on your own , without using pre-built code or library, for classifying a dataset. Perform necessary pre-processing steps. Analyze the importance of pre-processing. 11. **Probabilistic Supervised Learning - Naive Bayes:** Create a dataset from the sample given to you(e.g. “Play Tennis Probability”, “Shopper Buying Probability” etc.). Perform the necessary pre-processing steps such as encoding. Train the model using Naive Bayes Classifier. Give new test data and predict the classification output. Handcode the classification probability and compare with the model output. Analyze and write the inference. 12. **Supervised Learning Algorithms - Support Vector Machines:** Generate a separable dataset of size 1000 and 2 features. Plot the samples on a graph and mark the support vectors for the dataset. Also, show that changing the vectors other than the support vectors has no effect on the decision boundary. 13. **Supervised Learning Algorithms - Support Vector Machines:** Use SVM to classify the flowers in Iris dataset. Visualize the results for each of the following combinations:   **(a)** For every pair of (different) features in the dataset (there are 4). Which pair separates the data easily?  **(b)** Using One-vs-Rest and using One-vs-One. Which one fits better? Which one is easier to compute? Why?  **(c)** Using different kernels (Linear, RBF, Quadratic).   1. **Un-Supervised Learning Algorithms - Clustering:** Using any dataset from the UCI repository implement any one type of Hierarchical and Partitional Clustering you are familiar with. Plot the Dendrogram for Hierarchical Clustering and analyze your result. Plot the clustering output for the same dataset using these two partitioning techniques. Compare the results. Write the inference. 2. **Un-Supervised Learning Algorithms - K-Means Clustering:** Build a K-Means Model for the given dataset. In K-Means choosing the K value that gives a better model is always a challenge. We increase the value of K with a dataset having N points, the likelihood of the model increases, and obviously K<N, so to rank or maximize the likelihood we use BIC(Bayesian Information Criterion. Now,   **(a)** Build a K-Means Model for the given Dataset (You can use the library functions)  **(b)** Implement the BIC function that takes the cluster and data points and returns BIC value  **(c)** Implement a function to pick the best K value, that is maximize the BIC.  **(d)** Visualize the pattern found by plotting K v/s BIC. | |
| **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 and implement the various Machine Learning Algorithms in the realm of supervised and unsupervised learning. **(PO – 1(2),2(3),3(3),4(3),5(2),12(3)) & (PSO – 1(3), 2(2))** |
| **2.** | Demonstrate the working principle of these different ML models, determine their performance, usage and their applications. **(PO – 1(2), 3(2), 10(3)) & (PSO – 1(3),3(3))** |
| **3.** | Analyze the results and produce substantial written documentation. **(PO – 1(2),4(3),10(3),12(2)) & PSO – (1(2),2(2),3(2))** |

**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)