

Machine Learning lab		Semester	6
Course Code	BCSL606	CIE Marks	50
Teaching Hours/Week (L:T:P: S)	0:0:2:0	SEE Marks	50
Credits	01	Exam Hours	100
Examination type (SEE)	Practical		
Course objectives:			
1. To become familiar with data and visualize univariate, bivariate, and multivariate data using statistical techniques and dimensionality reduction.			
2. To understand various machine learning algorithms such as similarity-based learning, regression, decision trees, and clustering.			
3. To familiarize with learning theories, probability-based models and developing the skills required for decision-making in dynamic environments.			
Sl.NO	Experiments		
1	Develop a program to create histograms for all numerical features and analyze the distribution of each feature. Generate box plots for all numerical features and identify any outliers. Use California Housing dataset. Book 1: Chapter 2		
2	Develop a program to Compute the correlation matrix to understand the relationships between pairs of features. Visualize the correlation matrix using a heatmap to know which variables have strong positive/negative correlations. Create a pair plot to visualize pairwise relationships between features. Use California Housing dataset. Book 1: Chapter 2		
3	Develop a program to implement Principal Component Analysis (PCA) for reducing the dimensionality of the Iris dataset from 4 features to 2. Book 1: Chapter 2		
4	For a given set of training data examples stored in a .CSV file, implement and demonstrate the Find-S algorithm to output a description of the set of all hypotheses consistent with the training examples. Book 1: Chapter 3		
5	Develop a program to implement k-Nearest Neighbour algorithm to classify the randomly generated 100 values of x in the range of $[0,1]$. Perform the following based on dataset generated. <ol style="list-style-type: none"> 1. Label the first 50 points $\{x_1, \dots, x_{50}\}$ as follows: if $(x_i \leq 0.5)$, then $x_i \in \text{Class}_1$, else $x_i \in \text{Class}_2$ 2. Classify the remaining points, x_{51}, \dots, x_{100} using KNN. Perform this for $k=1,2,3,4,5,20,30$ Book 2: Chapter – 2		
6	Implement the non-parametric Locally Weighted Regression algorithm in order to fit data points. Select appropriate data set for your experiment and draw graphs Book 1: Chapter – 4		
7	Develop a program to demonstrate the working of Linear Regression and Polynomial Regression. Use Boston Housing Dataset for Linear Regression and Auto MPG Dataset (for vehicle fuel efficiency prediction) for Polynomial Regression. Book 1: Chapter – 5		
8	Develop a program to demonstrate the working of the decision tree algorithm. Use Breast Cancer Data set for building the decision tree and apply this knowledge to classify a new sample. Book 2: Chapter – 3		

9	Develop a program to implement the Naive Bayesian classifier considering Olivetti Face Data set for training. Compute the accuracy of the classifier, considering a few test data sets. Book 2: Chapter – 4
10	Develop a program to implement k-means clustering using Wisconsin Breast Cancer data set and visualize the clustering result. Book 2: Chapter – 4
<p>Course outcomes (Course Skill Set):</p> <p>At the end of the course the student will be able to:</p> <ol style="list-style-type: none"> 1. Illustrate the principles of multivariate data and apply dimensionality reduction techniques. 2. Demonstrate similarity-based learning methods and perform regression analysis. 3. Develop decision trees for classification and regression problems, and Bayesian models for probabilistic learning. 1. Implement the clustering algorithms to share computing resources. 	