B.Sc(H) Computer Science VI Sem
Practical List For Machine Learning

Question 1.

Classify the iris dataset using a decision tree classifier. Divide the dataset into training and testing in the ratio 80:20. Use the functions from the sklearn package. Display the final decision tree.

Question 2.

Classify the iris dataset using a Bayes classifier. Divide the dataset into training and testing in the ratio 80:20. Use the functions from the sklearn package. Assume the data follows a gaussian distribution. Display the training and testing accuracy, confusion matrix.

Question 3.

Classify the iris dataset using the KNN classifier. Divide the dataset into training, validation, and testing in the ratio 70:15:15. Use the functions from the sklearn package. Find the best value for k. Normalize the dataset before applying the model. Display the training, validation, and testing accuracy, confusion matrix.

Question 4.

Create a linear regression model using ordinary least squares estimation. Find the best fit line for the dataset 'salary.csv' using the above model. Display the training and testing dataset in the scatter plot and draw the best fit line in the same. Also find the MSE and R2 for the testing dataset.

Question 5

Consider the dataset california_housing from sklearn . Find the correlation b/w the different attributes of this dataset. Using the least square estimation method from sklearn, find the best fit line. Also find the error.

Question 6

Consider the dataset 'Adveristing.csv'. Find the correlation coefficient between the input attributes TV, Radio, Newspaper and Output Attribute Sales. Use least square estimation method to find the line of regression b/w
1. TV and Sales
2. Radio and Sales

3. Newspaper and Sales

For all of the above options, also draw a scatter plot and line of regression. Also find the error in each of the above.

Question 7

Consider the dataset 'Adveristing.csv'. Find the best fit regression line between the input attributes TV, Radio, Newspaper and Output Attribute Sales using gradient descent method. Also find R2 .

Question 8

Use logistic regression to build a model to classify the breast cancer dataset Divide the dataset into training and testing in the ratio 70:30 . Print the confusion matrix, sensitivity, specificity. For each iteration of training, store the training and testing accuracy. Plot a graph showing training and testing accuracy Vs iteration no. Do not use sklearn logistic function.

Question 9

Using logistic regression to build a model to classify the iris dataset. Divide the dataset into training and testing in the ratio 80:20 . Print the confusion matrix, sensitivity and specificity.

Question 10

Create a linear regression model using the gradient descent method. Create a class to represent the model with the following functions - init, fit and predict. Find the best fit line for the dataset Also find the MSE and R2 for the testing dataset.

Question 11

Consider the dataset wine from sklearn. Using PCA reduce the dimensionality of the dataset to 5. Build a classification model using gaussian naive bayes classifier. Find the training accuracy and test accuracy.

Question 12

Consider the dataset iris. Apply the PCA method to select the best 2 features. Using these features plot the scatter graph. Apply k-means clustering algorithm to cluster the transformed dataset into 3 clusters.

Question 13

Write a program to implement a single layer perceptron model. Train this for solving a AND problem with 3 variables.

Question 14

Consider the dataset iris. Apply hierarchical clustering algorithm to cluster the dataset into 3 clusters.

Question 15

Write a program to implement 2-layered ANN for classifying digits datasets from sklearn. Use 70% data for training the model and check the accuracy of the model on remaining 30% data. Use softmax activation function in the last layer and relu function in the hidden layer.