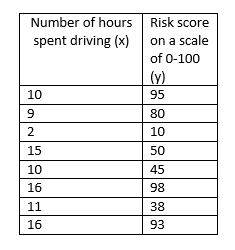
1.Assignment on Linear Regression:

The following table shows the results of a recently conducted study on the correlation of the

number of hours spent driving with the risk of developing acute backache. Find the equation of

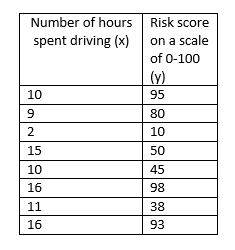
the best fit line for this data.



2.The following table shows the results of a recently conducted study on the correlation of the

number of hours spent driving with the risk of developing acute backache. Find the equation of

the curve using polynomial regression and demonstrate the degree comparison.



3. Implement multiple linear regression using n-features in the provided dataset. Evaluate the model using metrics like R2

Dataset- MultipleLinearRegression.csv

4.Assignment on Decision Tree Classifier:

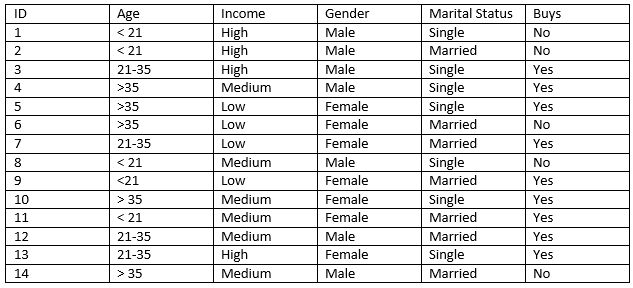
A dataset collected in a cosmetics shop showing details of customers and whether or not they

responded to a special offer to buy a new lip-stick is shown in table below. Use this dataset to build a decision tree, with Buys as the target variable, to help in buying lip-sticks in the future.

Find the root node of decision tree. According to the decision tree you have made from

previous training data set, what is the decision for the test data: [Age < 21, Income = Low,

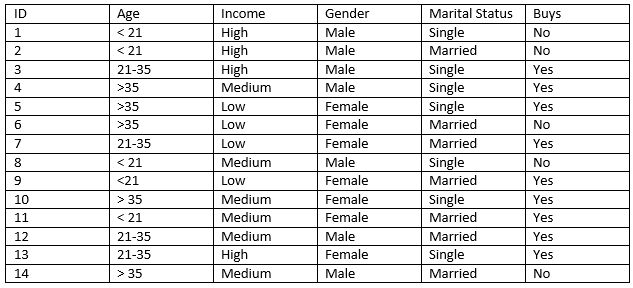
Gender = Female, Marital Status = Married]?



5 A dataset collected in a cosmetics shop showing details of customers and whether or not they

responded to a special offer to buy a new lip-stick is shown in table below. Use this dataset to build a random forest model, with Buys as the target variable, to help in buying lip-sticks in the future.

Compare the accuracy between decision tree and random forest model.

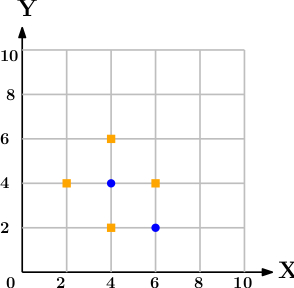


6. Assignment on k-NN Classification:

In the following diagram let blue circles indicate positive examples and orange squares indicate

negative examples. We want to use k-NN algorithm for classifying the points. If k=3, find the

class of the point (6,6). Extend the same example for Distance-Weighted k-NN and Locally

weighted Averaging

7.We have given a collection of 8 points. P1=[0.1,0.6] P2=[0.15,0.71] P3=[0.08,0.9] P4=[0.16,

0.85] P5=[0.2,0.3] P6=[0.25,0.5] P7=[0.24,0.1] P8=[0.3,0.2]. Perform the k-mean clustering

with initial centroids as

m1=P1 =Cluster#1=C1 and m2=P8=cluster#2=C2. Answer the

following

1] Which cluster does P6 belongs to?

2] What is the population of cluster around m2?

3] What is updated value of m1 and m2?

8.Apply Support Vector Machine algorithm for performing classification on iris dataset..

1]Classify the dataset into three species (Iris Setosa, Iris versicolor, Iris virginica)

2] Perform analysis and derive confusion matrix, classification report

Dataset-Iris.csv

9. Implement logistic regression on the pima indian diabetes dataset. Visualise the performance using ROC curve.

Dataset-Pimaindiandiabetes.csv

10.Given an example of weather conditions and playing sports. You need to calculate the probability of playing sports.Using naive bayes classifier classify whether players will play or not, based on the weather condition.

1) Derive confusion matrix

2)Find out the accuracy of the classifier

3)Illustrate with proper calculations.

Dataset-naivebayes.csv