1. Write a program to find the equation of the best fit line for the following data. Calculate R2 value.

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.



1. Write a program to find the equation of the best fit line for the following data. Create a scatter plot of the data with the estimated regression line.

The sales of a company (in million dollars) for each year are shown in the table below.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| x (year) | 2005 | 2006 | 2007 | 2008 | 2009 |
| y (sales) | 12 | 19 | 29 | 37 | 45 |

1. Write a program to find the least square regression line y = a x + b for following data.  
   and Estimate the value of y when x = 10.

The values of y and their corresponding values of y are shown in the table below

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| x | 0 | 1 | 2 | 3 | 4 |
| y | 2 | 3 | 5 | 4 | 6 |

1. Write a program to do : 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]?



1. Write a program to build decision tree for the following data.



1. Write a program to do following:

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?

1. Write a program to build decision tree for following data.

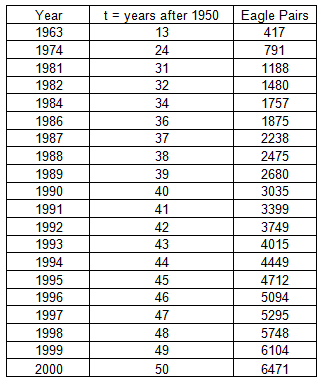


1. In the following diagram let blue circles indicate positive examples and orange squares indicate negative examples. We want to use k-NN algorithm for classifying the points. If k=3, find the class of the point (6,6).



1. Write a program on following data to

a) Find the least square regression line y = a x + b.



10. Write a program on following data to

a) Find the least square regression line y = a x + b.   
 b) Use the least squares regression line as a model to estimate the sales of the company in 2012.

The sales of a company (in million dollars) for each year are shown in the table below.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| x (year) | 2005 | 2006 | 2007 | 2008 | 2009 |
| y (sales) | 12 | 19 | 29 | 37 | 45 |

1. Write a program to apply k-NN on following data and find out the T Shirt Size on New customer named 'Monica' has height 161cm and weight 61kg.

|  |  |  |
| --- | --- | --- |
| **Height (in cms)** | **Weight (in kgs)** | **T Shirt Size** |
| 158 | 58 | M |
| 158 | 59 | M |
| 158 | 63 | M |
| 160 | 59 | M |
| 160 | 60 | M |
| 163 | 60 | M |
| 163 | 61 | M |
| 160 | 64 | L |
| 163 | 64 | L |
| 165 | 61 | L |
| 165 | 62 | L |
| 165 | 65 | L |
| 168 | 62 | L |
| 168 | 63 | L |
| 168 | 66 | L |
| 170 | 63 | L |
| 170 | 64 | L |
| 170 | 68 | L |