

1. Find the equation of the best fit line for the following data. Calculate R2 value and draw ROC curve
2. Find the equation of the best fit line for the following data. Create a scatter plot of the data with the estimated regression line.
3. 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]? Calculate entropy and Information gain
4. Calculate entropy and Information gain for the following data

price	maintenance	capacity	airbag	profitable
low	low	2	no	yes
low	med	4	yes	no
low	low	4	no	yes
low	high	4	no	no
med	med	4	no	no
med	med	4	yes	yes
med	high	2	yes	no
med	high	5	no	yes
high	med	4	yes	yes
high	high	2	yes	no
high	high	5	yes	yes

5. 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?

6. Draw decision tree for following data

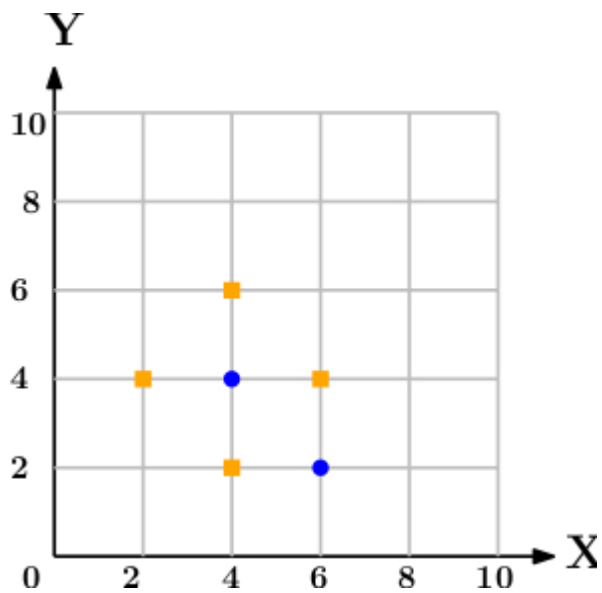
outlook	temperature	humidity	wind	playtennis
sunny	hot	high	weak	no
sunny	hot	high	strong	no
overcast		hot	high	weak
rain	mild	high	weak	yes
rain	cool	normal	weak	yes
rain	cool	normal	strong	no
overcast		cool	normal	strong
sunny	mild	high	weak	no
sunny	cool	normal	weak	yes

rain	mild	normal	weak	yes
sunny	mild	normal	strong	yes
overcast		mild	high	strong yes
overcast		hot	normal	weak yes
rain	mild	high	strong	no

7. Apply KNN on IRIS data set having 4 attributes and 1 class label

5.1,3.5,1.4,0.2,Iris-setosa
 4.9,3.0,1.4,0.2,Iris-setosa
 4.7,3.2,1.3,0.2,Iris-setosa
 4.6,3.1,1.5,0.2,Iris-setosa
 5.0,3.6,1.4,0.2,Iris-setosa
 5.4,3.9,1.7,0.4,Iris-setosa

KNN



Linear regression problem set

Year	t = years after 1950	Eagle Pairs
1963	13	417
1974	24	791
1981	31	1188
1982	32	1480
1984	34	1757
1986	36	1875
1987	37	2238
1988	38	2475
1989	39	2680
1990	40	3035
1991	41	3399
1992	42	3749
1993	43	4015
1994	44	4449
1995	45	4712
1996	46	5094
1997	47	5295
1998	48	5748
1999	49	6104
2000	50	6471

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

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

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

New customer named 'Monica' has height 161cm and weight 61kg. apply KNN