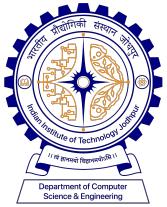
शिक्षा मंत्रालय
MINISTRY OF
EDUCATION
सत्यमेव जयते

INDIAN INSTITUTE OF TECHNOLOGY
JODHPUR

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

NPTEL



PMRF
Prime Minister's Research Fellowship

Week 4 - Live Session

Data Mining

Swapnil S. Mane
mane.1@iitj.ac.in
PMRF Research Scholar

Q1-3 are based on a simple Bayesian Network shown below:

P(x=1)=0.60	P(y=1|x=1)=0.40	P(z=1|y=1)=0.25	P(w=1|z=1)=0.45
	P(y=1|x=0)=0.30	P(z=1|y=0)=0.60	P(w=1|z=0)=0.30

The Bayesian Network is fully specified by the marginal probabilities of the root node(x) and the conditional probabilities.

Q1. P(y=0) is:

a) 0.70
b) 0.12
c) 0.64
d) 0.36

$P(y=0) = 1 - P(y=1)$ (1)

$P(y=1) = P(y=1|x=1) * P(x=1) + P(y=1|x=0) * P(x=0)$

$P(y=1) = (0.40 * 0.60) + (0.30 * 0.40)$

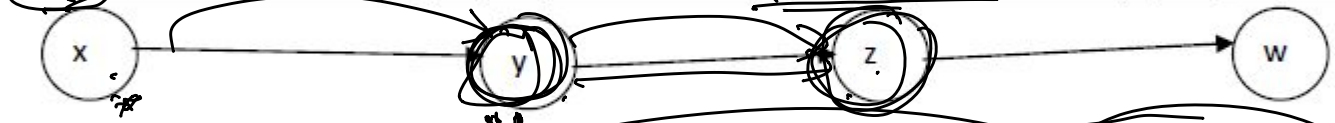
$P(y=1) = \frac{0.36}{0.36}$

$P(y=0) = 1 - 0.36$

$= 0.64$

Q1-3 are based on a simple Bayesian Network shown below:

P(x=1)=0.60	P(y=1|x=1)=0.40	P(z=1|y=1)=0.25	P(w=1|z=1)=0.45
	P(y=1|x=0)=0.30	P(z=1|y=0)=0.60	P(w=1|z=0)=0.30



The Bayesian Network is fully specified by the marginal probabilities of the root node(x) and the conditional probabilities.

$P(y=1|x=1) + P(y=0|x=1) = 1$

Q2. P(z=1|x=1) is:

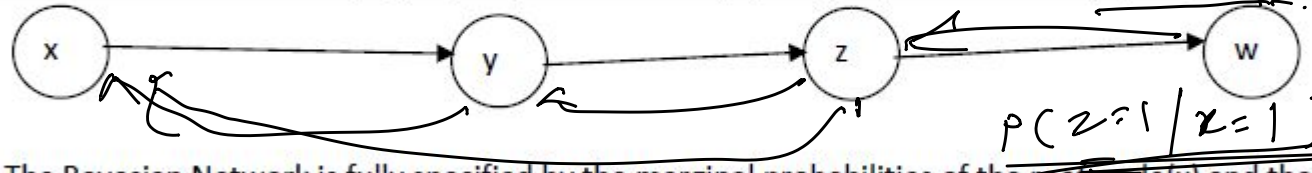
a) 0.50
b) 0.60
c) 0.46
d) 0

$P(z=1|x=1) = P(z=1|y=1) * P(y=1|x=1) + P(z=1|y=0) * P(y=0|x=1)$

$= 0.46$

Q1-3 are based on a simple Bayesian Network shown below:

P(x=1)=0.60	P(y=1|x=1)=0.40 P(y=1|x=0)=0.30	P(z=1|y=1)=0.25 P(z=1|y=0)=0.60	P(w=1|z=1)=0.45 P(w=1|z=0)=0.30



The Bayesian Network is fully specified by the marginal probabilities of the root node(x) and the conditional probabilities.

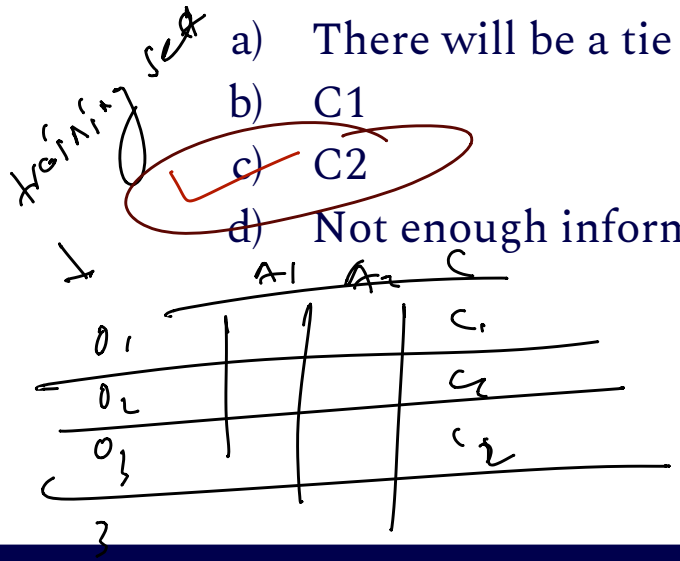
Q3. P(w=0|x=1) is:

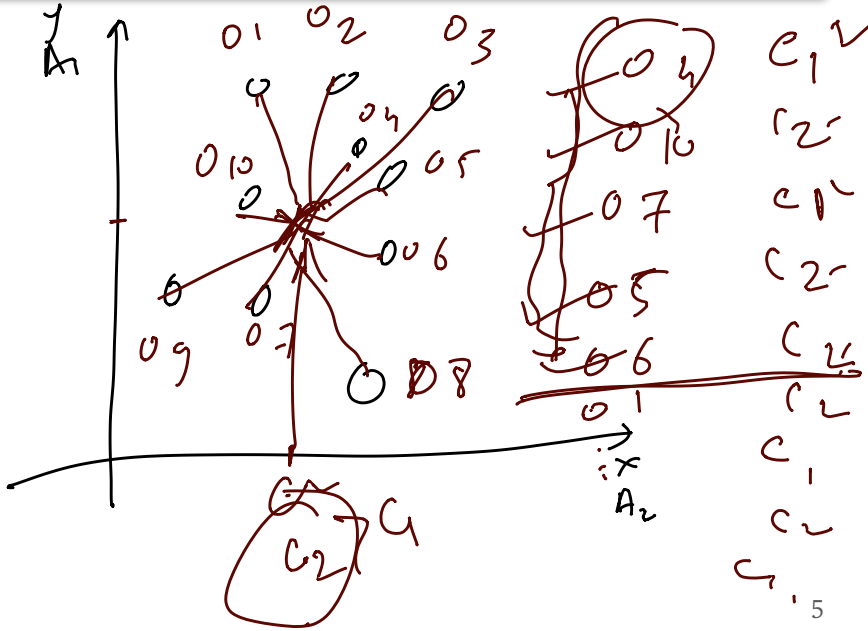
a) 0.37
b) 0.63
c) 1
d) None of the above

$$P(w=0|x=1) = P(w=0|z=1) * P(z=1|x=1) + P(w=0|z=0) * P(z=0|x=1)$$
$$= (0.55 * 0.46) + (0.70 * 0.54)$$
$$= 0.63$$

Q4. Consider a binary classification problem with two classes C1 and C2. Class labels of ten other training set instances sorted in increasing order of their distance to an instance x is as follows: C1, C2, C1, C2, C2, C2, C1, C2, C1, C2. How will a K=5 nearest neighbor classifier classify x?

a) There will be a tie
b) C1
c) C2
d) Not enough information to classify





Consider the following data for questions 5-6.

You are given the following set of training examples. Each attribute can take value either 0 or 1.

A1	A2	A3	Class
0	0	1	C1
0	1	0	C1
0	1	1	C1
1	0	0	C2
1	1	0	C1
1	1	1	C2

Q5. How would a 3-NN classify the example A1 = 1, A2 = 0, A3 = 1 if the distance metric is Euclidean distance?

a) C1

b) C2

c) There will be a tie

d) Not enough information to classify

Consider the following data for questions 5-6.

You are given the following set of training examples. Each attribute can take value either 0 or 1.

A1	A2	A3	Class
0	0	1	C1
0	1	0	C1
0	1	1	C1
1	0	0	C2
1	1	0	C1
1	1	1	C2

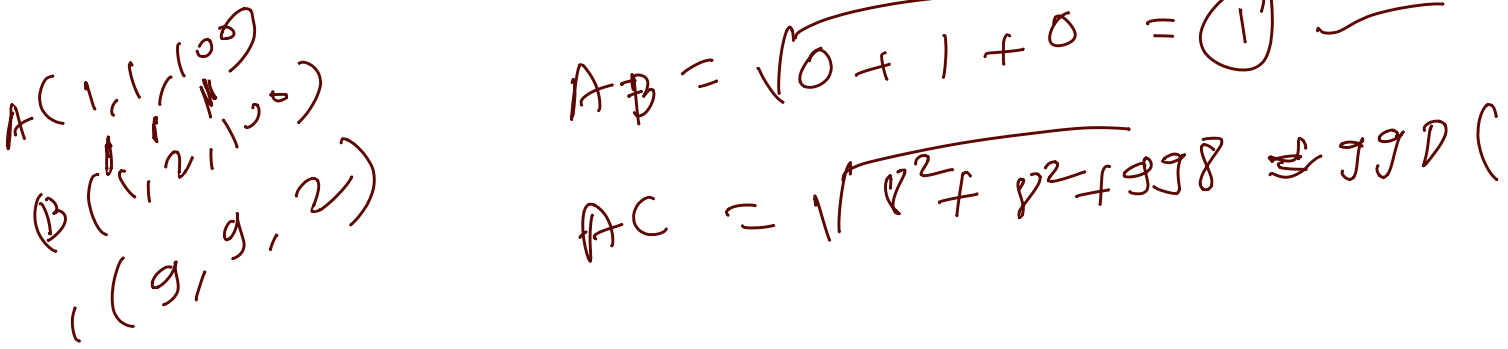
Q6. How would a 3-NN classify the example A1 = 0, A2 = 0, A3 = 0 if the distance metric is Euclidean distance?

a) C1

b) C2

c) There will be a tie

d) Not enough information to classify


A(1,1,100)
B(1,2,1,0)
(9,9,2)
AB = √0+1+0 = 1
AC = √9²+9²+998 ≈ 999

Q7. Issues with Euclidean measure are:

a) High dimensional data.
b) Can produce counter-intuitive results.
c) Shrinking density – sparsification effect
d) All of the above.