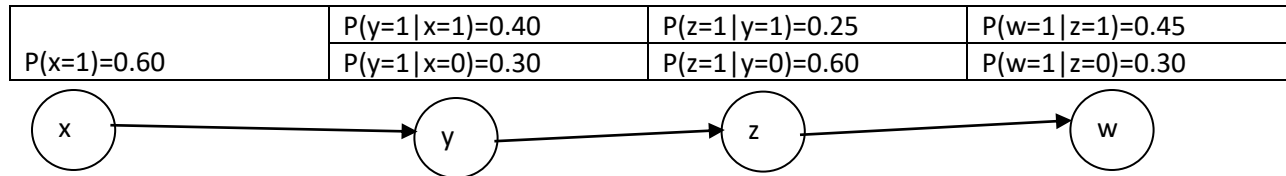## Assignment Week 4: Bayes Classifier and KNN(2024)

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.

1. P(y=0) is: **(2 marks)**

A. 0.70

B. 0.12

C. 0.64

D. 0.36

Ans: C

Explanation: P(y=0) = 1-P(y=1)

 P(y=1) = P(y=1|x=0)*P(x=0) + P(y=1|x=1)*P(x=1) = 0.30*0.40 + 0.40*0.60 = 0.36

P(y=0) = 1-0.36 = 0.64

2. P(z=1|x=1) is: **(2 marks)**

A. 0.50

B. 0.60

C. 0.46

D. 0

Ans: C

Explanation:

 P(z=1|x=1) = P(z=1|y=0)*P(y=0|x=1) + P(z=1|y=1)*P(y=1|x=1) = 0.60*0.60 + 0.25*0.40 = 0.46

3. P(w=0|x=1) is: **(2 marks)**

A. 0.37

B. 0.63

C. 1

D. None of the above

Ans: B

Explanation: 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)

$$= 0.55*0.46 + 0.70*0.54 = 0.63$$

4. 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*? **(1 mark)**

A. There will be a tie

B. C1

C. C2

D. Not enough information to classify

Ans: C

Explanation: closest 3 neighbors are C1, C2, C1, C2, C2. In this C1 has 2 occurrences and C2 has 3 occurrence, therefore, by majority voting X will be classified as C2.

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

5. How would a 3-NN classify the example *A1* = 1, *A2* = 0, *A3* = 1 if the distance metric is Euclidean distance? **(1 mark)**

A. C1

B. C2

C. There will be a tie

D. Not enough information to classify

Ans: B

Explanation: We get minimum distance of 1 with points (0,0,1), (1,0,0), (1,1,1) which are classified as C1, C2, C2; since majority is C2 therefore class is C2.

6. How would a 3-NN classify the example $A_1 = 0$, $A_2 = 0$, $A_3 = 0$ if the distance metric is Euclidean distance? **(1 mark)**

A. C1

B. C2

C. There will be a tie

D. Not enough information to classify

Ans: A

Explanation: We get minimum distance of 1 with points (0,0,1), (0,1,0), (1,0,0) which are classified as C1, C1, C2; since majority is C1 therefore class is C1.

7. Issues with Euclidean measure are: **(1 mark)**

A. High dimensional data.

B. Can produce counter-intuitive results.

C. Shrinking density – sparsification effect

D. All of the above.

Ans: D

Explanation: all the above are issues of Euclidean measure.