ML

LECTURE-13

BY
Dr. Ramesh Kumar Thakur
Assistant Professor (II)
School Of Computer Engineering

KNN Previous Year Questions

- 1.Using KNN algorithm and the given data set, predict the label of the test data point (3,7), where K=3 and Euclidean distance.

X	Y	Label
7	7	1
7	4	1
3	4	2
1	4	2

- Ans :- To predict the label of the test data point, we have to calculate the distance from the test data point to other data in the data set using the Euclidean distance formula.

- **Euclidean distance formula : $\sqrt{(X_2-X_1)^2+(Y_2-Y_1)^2}$**
- Where:
- X_2 = Test data point's X value (3).
- X_1= Existing data's X value.
- Y_2 = Test data point's Y value (7).
- Y_1 = Existing data's Y value.

KNN Previous Year Questions

- **Distance #1**
- For the first row, d1:
- $d1 = \sqrt{(3 - 7)^2 + (7 - 7)^2}$
- $= \sqrt{16 + 0}$
- $= \sqrt{16}$
- $= 4$

- **Distance #2**
- For the second row, d2:
- $d2 = \sqrt{(3 - 7)^2 + (7 - 4)^2}$
- $= \sqrt{16 + 9}$
- $= \sqrt{25}$
- $= 5$

KNN Previous Year Questions

- **Distance #3**
- For the third row, d3:
- $d3 = \sqrt{(3 - 3)^2 + (7 - 4)^2}$
- $= \sqrt{0 + 9}$
- $= \sqrt{9}$
- $= 3$

- **Distance #4**
- For the fourth row, d4:
- $d4 = \sqrt{(3 - 1)^2 + (7 - 4)^2}$
- $= \sqrt{4 + 9}$
- $= \sqrt{13}$
- $= 3.6$

KNN Previous Year Questions

- Here's what the table will look like after all the distances have been calculated:

X	Y	Label	Distance
7	7	1	4
7	4	1	5
3	4	2	3
1	4	2	3.6

- As we can see, the **majority class within the 3 nearest neighbors to the test data point is label 2. Therefore, we'll classify the test data point as label 2.**

- 2.Perform KNN classification on the following training instances each having two attributes (X1, and X2). Compute the class label for the test instance t1 = (3,7), with K=3 and Euclidean distance.

Training instances	X1	X2	output
I1	7	7	0
I2	7	4	0
I3	3	4	1
I4	1	4	1

- Ans :- Same as solution of question 1.

KNN Previous Year Questions

- 3.Explain the merits and demerits of Cosine distance measure. Find the cosine distance between (1,6,1,0) and (0,1,2,2).
- Ans:- Merits of cosine distance measure:-
- a) Low storage cost,
- b) High computational efficiency
- c) Good retrieval performance.
- Demerits of cosine distance measure:-
- I. It yields the same value regardless of the size of the vectors being compared, as long as the angle between them is the same.
- II. It does not take into account the semantic meanings of words or phrases, even when using techniques like Natural Language Processing.
- **Formula for cosine distance:-**

$$D_c = \frac{p.q}{||p||||q||}$$

- Where,
- p . q = product (dot) of the vectors ‘p’ and ‘q’.
- ||p|| and ||q|| = length (magnitude) of the two vectors ‘p’ and ‘q’.

KNN Previous Year Questions

- Cosine distance between (1,6,1,0) and (0,1,2,2) is

$$D_c = \frac{1\times 0 + 6\times 1 + 1\times 2 + 0\times 2}{\sqrt{1^2+6^2+1^2+0^2}\times\sqrt{0^2+1^2+2^2+2^2}} = \frac{8}{\sqrt{38}\times\sqrt{9}} = \frac{8}{18.49} = 0.43$$

- 4.Under what conditions Minkowski distance is same as Euclidean distance?
- Ans:- When the **order in the formula of Minkowski distance is 2, it is same as Euclidean Distance.**

- 5.Suppose you have a dataset of animals and you want to use KNN to predict whether a new animal is a cat or a dog based on its weight and height. You have the following dataset.

Animal	Weight (Kg)	Height (Cm)	Species
1	4	35	Cat
2	6	40	Dog
3	3	25	Cat
4	7	45	Dog
5	5	30	Cat
6	8	50	Dog
7	2	20	Cat
8	5	35	Dog

- Predict the species of a new animal that weights 4Kg and is 30 Cm tall.
- Ans: To predict the species of a new animal, we have to calculate the distance of the features of new animal from the features of other animal in the data set using the Euclidean distance formula.

KNN Previous Year Questions

- **Euclidean distance formula : $\sqrt{(X_2-X_1)^2+(Y_2-Y_1)^2}$**
- Where:
- X_2 = New animal's weight (4).
- X_1= Existing animal's weight.
- Y_2 = New animal's height (30).
- Y_1 = Existing animal's height.

- **Distance #1**
- For the first row, d1:
- $d1 = \sqrt{(4 - 4)^2 + (30 - 35)^2}$
- $= \sqrt{0 + 25}$
- $= \sqrt{25}$
- $= 5$

KNN Previous Year Questions

- **Distance #2**
- For the second row, d2:
- $d2 = \sqrt{(4 - 6)^2 + (30 - 40)^2}$
- $= \sqrt{4 + 100}$
- $= \sqrt{104}$
- $= 10.2$

- **Distance #3**
- For the third row, d3:
- $d3 = \sqrt{(4 - 3)^2 + (30 - 25)^2}$
- $= \sqrt{1 + 25}$
- $= \sqrt{26}$
- $= 5.1$

KNN Previous Year Questions

- **Distance #4**
- For the fourth row, d4:
- d4 = $\sqrt{(4 - 7)^2 + (30 - 45)^2}$
- = $\sqrt{9 + 225}$
- = $\sqrt{234}$
- = 15.3

- **Distance #5**
- For the first row, d5:
- d1 = $\sqrt{(4 - 5)^2 + (30 - 30)^2}$
- = $\sqrt{1 + 0}$
- = $\sqrt{1}$
- = 1

KNN Previous Year Questions

- **Distance #6**
- For the second row, d2:
- $d2 = \sqrt{(4 - 8)^2 + (30 - 50)^2}$
- $= \sqrt{16 + 400}$
- $= \sqrt{416}$
- $= 20.4$

- **Distance #7**
- For the third row, d3:
- $d3 = \sqrt{(4 - 2)^2 + (30 - 20)^2}$
- $= \sqrt{4 + 100}$
- $= \sqrt{104}$
- $= 10.2$

KNN Previous Year Questions

- **Distance #8**
- For the fourth row, d4:
- $d4 = \sqrt{(4 - 5)^2 + (30 - 35)^2}$
- $= \sqrt{1 + 25}$
- $= \sqrt{26}$
- $= 5.1$
- Here's what the table will look like after all the distances have been calculated:

Animal	Weight (Kg)	Height (Cm)	Species	Distance
1	4	35	Cat	5
2	6	40	Dog	10.2
3	3	25	Cat	5.1
4	7	45	Dog	15.3
5	5	30	Cat	1
6	8	50	Dog	20.4
7	2	20	Cat	10.2
8	5	35	Dog	5.1

- As we can see, the **majority class within the 3 nearest neighbors to the new animal is cat. Therefore, we'll classify the new animal as cat.**

KNN Previous Year Questions

- 6.Evaluate the Euclidean distance, Manhattan distance, Minkowski distance and the Cosine distance for the following two points. P1(1,0,2,5,3) and P2(2,1,0,3,-1).
- Ans:- **Euclidean distance formula:-**

$$D_e = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2}$$

- Euclidean distance between P1(1,0,2,5,3) and P2(2,1,0,3,-1) is

$$D_e = \sqrt{(1-2)^2 + (0-1)^2 + (2-0)^2 + (5-3)^2 + (3-(-1))^2} = \sqrt{26} = 5.1$$

- **Manhattan distance formula:-**

$$D_m = \sum_{i=1}^{n} |p_i - q_i|$$

- Manhattan distance between P1(1,0,2,5,3) and P2(2,1,0,3,-1) is

$$D_m = |1-2| + |0-1| + |2-0| + |5-3| + |3-(-1)| = 10$$

- **Minkowski distance formula:-**

$$D_{Minkowski} = \left(\sum_{i=1}^{n} |p_i - q_i|^k\right)^{\frac{1}{k}}$$

KNN Previous Year Questions

- Minkowski Distance is the **generalized form of Euclidean and Manhattan Distance. Here, k represents the order of the norm. When the order(k) is 1, it will represent Manhattan Distance and when the order in the above formula is 2, it will represent Euclidean Distance.** So Minkowski distance is10 when k=1 and Minkowski distance is 5.1 when k=2.
- **Cosine distance formula:-**

$$D_c = \frac{p.q}{\|p\|\|q\|}$$

- Where,
- p . q = product (dot) of the vectors ‘p’ and ‘q’.
- ||p|| and ||q|| = length (magnitude) of the two vectors ‘p’ and ‘q’.
- Cosine distance between P1(1,0,2,5,3) and P2(2,1,0,3,-1) is

$$D_c = \frac{1\times 2 + 0\times 1 + 2\times 0 + 5\times 3 + 3\times -1}{\sqrt{1^2+0^2+2^2+5^2+3^2}\times\sqrt{2^2+1^2+0^2+3^2+-1^2}} = \frac{14}{\sqrt{39}\times\sqrt{15}}$$

$$= \frac{14}{24.18} = 0.58$$

- 7.Why KNN is called as Lazy Learner algorithm?
- Ans:- KNN is called a lazy learner algorithm because **it does not learn from the training set immediately, instead it stores the dataset, and at the time of classification, it performs an action** on the dataset.

KNN Previous Year Questions

- 8.Perform KNN classification on the following dataset and predict the class for (height = 170, weight = 57), with K=5 using Euclidean distance.

Height (CM)	Weight(KG)	Class
167	51	Underweight
182	62	Normal
176	69	Normal
172	65	Normal
173	64	Normal
174	56	Underweight
169	58	Normal
173	57	Normal
170	55	Normal
170	57	?

- 9.Using KNN algorithm and the given data set, predict the label of the test data point (8,5), where K=3 and Euclidean distance.

X	Y	Label
4.2	3.8	0
6.5	7.7	1
7.3	8.6	1
5.7	5.9	0
8.0	8.1	1
10.0	6.5	1

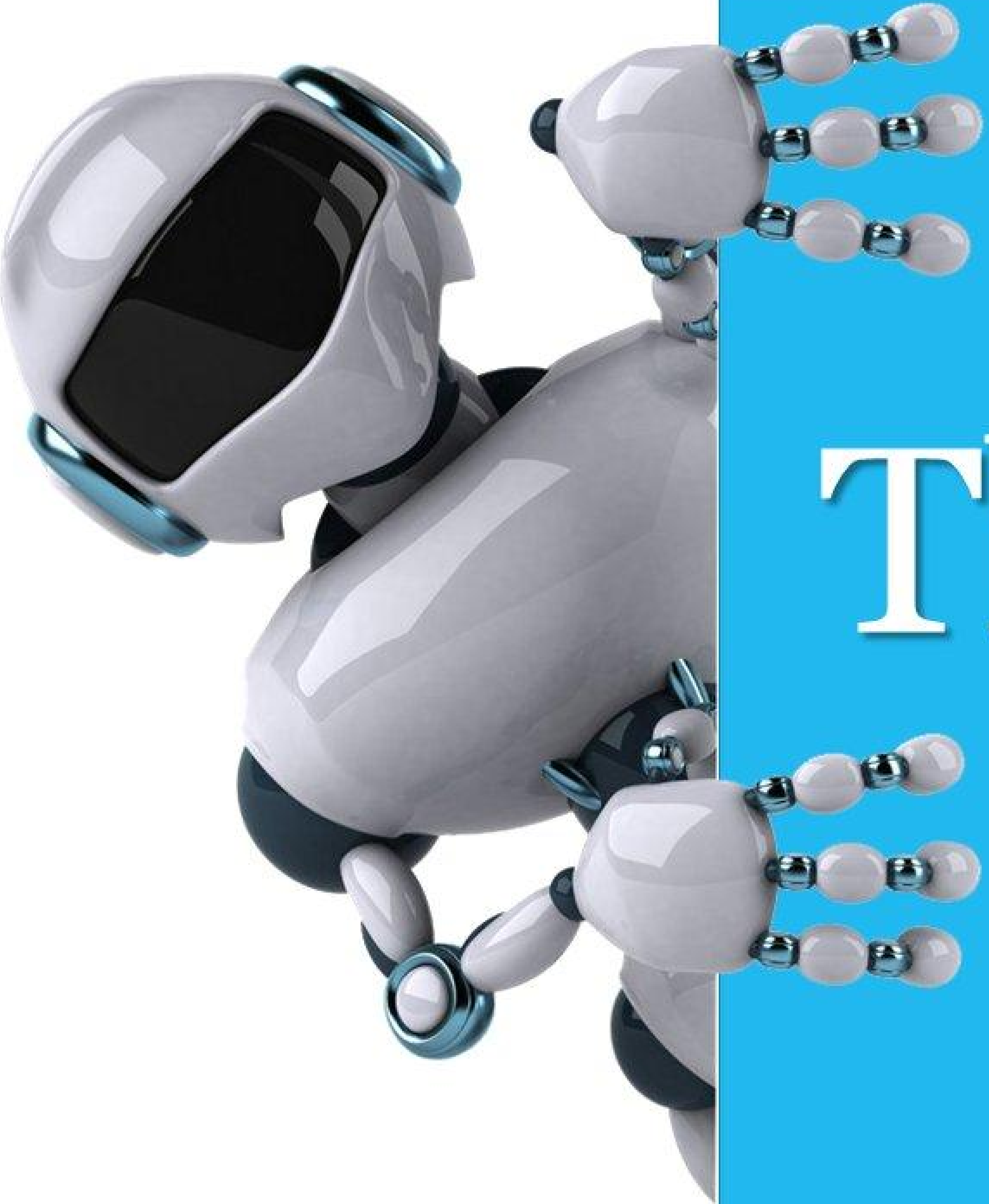
- 10.The Manhattan distance between two points (10, 10) and (30, 30) is ?

KNN Previous Year Questions

- 11.Using KNN algorithm and the given data set, predict the class label of the test data point (16,8), where K=3 and Euclidean distance.

X	Y	Label
10	5	0
6.5	11	1
7	15	1
12	5	0
8	10	1
15	8	0

- Note:- Question 8-11 is homework.



Thank you