Condition Monitoring of Structures, Machines and Processes

Tutorial 07

Voice: Dhanalekshmi

7

CONCEPT: CLASSIFICATION (K-NN)

k-Nearest Neighbour

What is k-NN?

- K-NN is one of the simplest machine learning algorithm based on supervised learning approach.
- K-NN algorithm assumes the similarity between the new data and the available data and assigns the new data into the category/class that is most similar to the available categories.
- K-NN algorithm stores all the available data and classifies a new data point based on the similarity.
- Hence, when a new data appears then it can be easily classified into well suited class by using k-NN algorithm

k-Nearest Neighbour

How k-NN works?

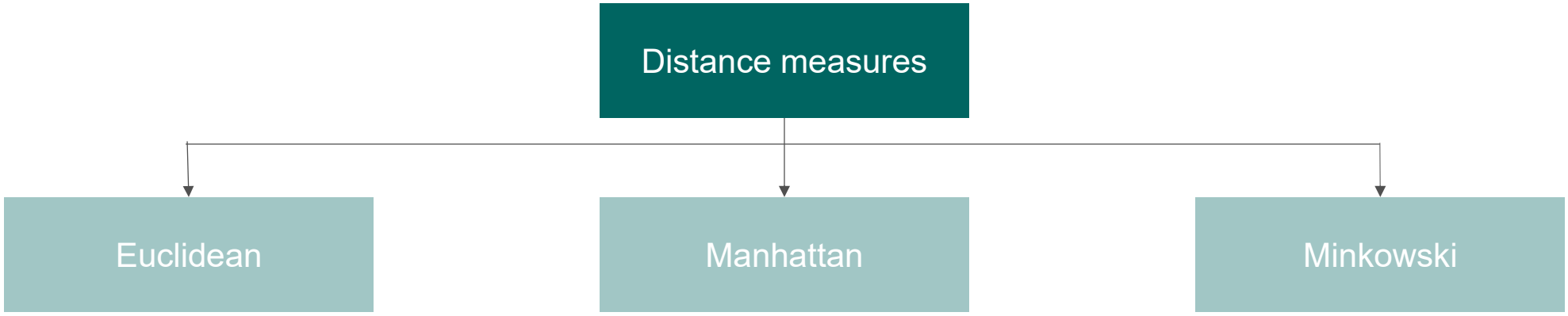
i. Select k number of neighbours

ii. Calculate the distance of k number of neighbours

iii. Take the k nearest neighbours as per the calculated distance measure.

iv. Among the k neighbours, count the number of data points in each class

v. Assign the new data points to that class for which the number of neighbour of the neighbour is maximum.



k-Nearest Neighbour

Example

- Perform k-NN algorithm on the following dataset and predict the class for the input X: p1=3 and p2=7.
- Consider k=3

p1	p2	Label/ Class
7	7	False
7	4	False
3	4	True
1	4	True

k-Nearest Neighbour

Example

i. Given number of neighbours, k=3

ii. Calculate the distance of k number of neighbours

- Here, Euclidean distance (d) measure is used

$$\boldsymbol{d = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2}}$$

x_1, x_2 is the value to be predicted
y_1, y_2 is the input values

k-Nearest Neighbour

Example

p1	p2	Distance	
7	7	$\sqrt{(3-7)^2+(7-7)^2}$	4
7	4	$\sqrt{(3-7)^2+(7-4)^2}$	5
3	4	$\sqrt{(3-3)^2+(7-4)^2}$	3
1	4	$\sqrt{(3-1)^2+(7-4)^2}$	3.6

- Take the minimum Euclidean distance
- From this table the minimum value is 3. This means that we need to take 3 Euclidean distances which are minimum
- These are d = 3, 3.6 and 4
- Lets name them as C1, C2 and C3 repectively

k-Nearest Neighbour

Example

p1	p2	Distance		Label/ Class	
7	7	$\sqrt{(3-7)^2+(7-7)^2}$	4	False	C3
7	4	$\sqrt{(3-7)^2+(7-4)^2}$	5	False	
3	4	$\sqrt{(3-3)^2+(7-4)^2}$	3	True	C1
1	4	$\sqrt{(3-1)^2+(7-4)^2}$	3.6	True	C2

- Now, we can say that C1 belongs to the True class, C2 belongs to the C2 class and C3 belongs to the False class.
- So, this way we have 2 True classes and one False class.
- Hence the majority goes to the True class.
- Therefore, X: p1=3 and p2=7 will belong to the class: True