K-NN Solved Numerical Example :⇒

Given a dataset as shown below :-

"Sepal length"	"Sepal_width"	"Species"
5.3	3.7	Setosa
5.1	3.8	Setosa
7.2	3.0	Virginica
5.4	3.4	Setosa
5.1	3.3	Setosa
5.4	3.9	Setosa
7.4	2.8	Virginica
6.1	2.8	Versicolor.
7.3	2.9	Virginica
6.0	2.7	Versicolor
5.8	2.8	Virginica
6.3	2.3	Versicolor
5.1	2.5	Versicolor
6.3	2.5	Versicolor
5.5	2.4	Versicolor.

"Target-Variable"

Now we are given a new example of a flower species & we need to determine its class given sepal length & sepal width

Test Example :⇒

Sepal length	Sepal width	Species
5.2	3.1	?

Step① :⇒ Find the distance :⇒

We need to compute the distance of the test example from all the training examples.

To compute the distance we can use either Euclidean distance or "Manhattan distance". or any other

So, here let us use Euclidean distance.

$$\text{Distance (Sepal length ; Sepal width)} = \sqrt{(x-a)^2 + (y-b)^2}$$

$$= \sqrt{(5.2-5.3)^2 + (3.1-3.7)^2}$$

Distance (Sepal length, Sepal width) = 0.608.

[Here 'x & y' are Sepal length & Sepal width of test example & 'a' & 'b' are the Sepal length & Sepal width of each of the training samples]

Sepal Length	Sepal width	Species	Distance
5.3	3.7	Setosa	0.608

In this way we need to compute the Euclidean distance of test example from all the training samples.

P.T.O

Computed distance :⇒

Sepal length	Sepal width	Species	Distance	Rank
5.3	3.7	Setosa	0.608	3
5.1	3.8	Setosa	0.707	6
7.2	3.0	Virginica	2.002	13
5.4	3.4	Setosa	0.36	2
5.1	3.3	Setosa	0.22	1
5.4	3.9	Setosa	0.82	8
7.4	2.8	Virginica	2.22	15
6.1	2.8	Versicolor	0.94	10
7.3	2.9	Virginica	2.1	14
6.0	2.7	Versicolor	0.89	9
5.8	2.8	Virginica	0.67	5
6.3	2.3	Versicolor	1.36	12
5.1	2.5	Versicolor	0.60	4
6.3	2.5	Versicolor	1.25	11
5.5	2.4	Versicolor	0.75	7

Step No. 2 :⇒ Compute the Ranks.

We need to compute the ranks of new example wrt the given examples, i.e. the distance which is min. will be having the first rank, the distance which is 2nd minimum will be having the second rank & so-on.

In this case '0.22' is the smallest, so it will be having the first rank. Next smallest is 0.36, it will be given 2nd rank & so-on

∴ we need to sort the things in accendig oder on the basis of distance metric, we just computed.

Step no. 3 :> Given the value of 'K', find the nearest neighbour.

If K = 1 then it is setosa

If K = 2 (Both are saying it is setosa)

∴ the test point is setosa.

if K = 5 , Here we need to find out first five nearest neighbour

1 → setosa 4 → versicolor
2 → setosa 5 → virginica
3 → setosa

∴ 3 are setosa, hence test example belongs to setosa
