Page No.
Date

K - Nearest Neighbor Algorithm

step - by step on how to compute K-nearest neighbor algorithm.

1. Determine parameter K = number of nearest neighbor.
2. Calculate the distance between the query-instance and all the training samples.
3. sort the distance and determine nearest neighbors based on the K-th minimum distance.
4. Gather the category Y of the nearest neighbors
5. Use simple majority of the category of nearest neighbors as the prediction value of the query instance

Ques We have data from the questionnaires survey (to ask people opinion) and objective testing with two attributes (acid durability and strength) to classify whether a special paper tissue is good or not. Here are four training samples

Acid durability X_1	Strength X_2	classification Y
7	7	Bad
7	4	Good
3	4	Good
1	4	Good

Page No.
Date

Now the factory produces a new paper tissue that pass laboratory test with

$X1 = 3$

$X2 = 7$

Without another expensive survey, can we guess what the classification of this new tissue is?

Prediction

1. Determine parameter, K = number of nearest neighbor.

use $K = 3$.

2. Calculate the distance b/w the query instance and all the training samples.

Coordinate for query instance is (3, 7). [$\because X1 = 3, X2 = 7$]

instead of calculating distance we compute square distance which is faster to calculate. (without square root).

X1	X2	Distance to query instance (3,7)
7	7	$(7-3)^2 + (7-7)^2 = 16$
7	4	$(7-3)^2 + (4-7)^2 = 25$
3	4	$(3-3)^2 + (4-7)^2 = 9$
1	4	$(1-3)^2 + (4-7)^2 = 13$

Page No.
Date

3 Sort the distance and determine nearest neighbor based on the K th - minimum distance.

X_1	X_2	Square distance to query instance (3, 7)	Rank min. distance	Is it included in 3-Nearest Neighbor
7	7	16	3	Yes
7	4	25	4	NO
3	4	9	1	Yes
1	4	13	2	Yes.

4 Gather the category Y of the nearest neighbor. Notice in the second row last column that the category of nearest neighbor (Y) is not included because the rank of this data is more than 3. (K = 3)

X_1	X_2	Distance of query instance (3, 7)	Rank min. distance	Is it included in 3-nearest neighbor	Y = category of nearest neighbor.
7	7	16	3	Yes	Bad
7	4	25	4	NO	—
3	4	9	1	Yes	Good
1	4	13	2	Yes	Good

Page No.
Date

5 Use simple majority majority of the category of nearest neighbors as the prediction value of the query instance.

We have 2 good and 1 bad, since 2>1, we conclude that a new paper tissue with X1 = 3, & X2 = 7 is included in Good category.