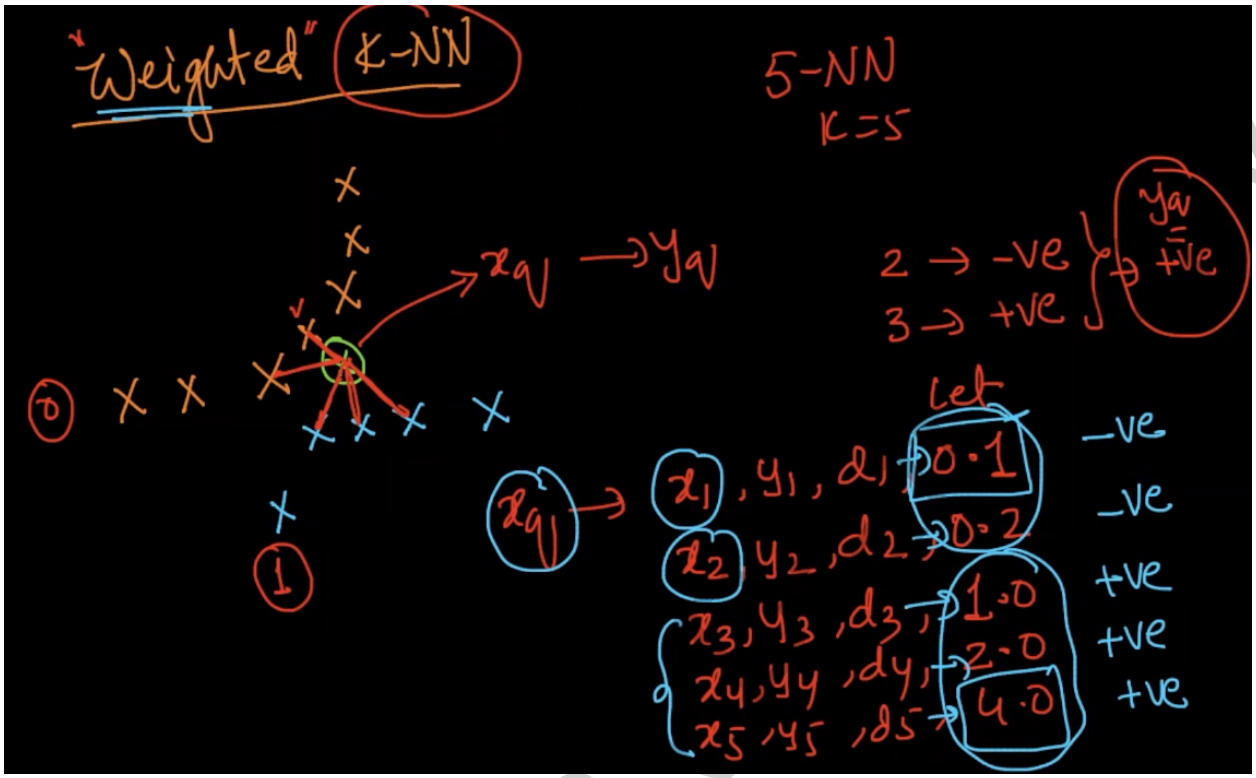
29.19 Weighted K-NN



So far we have seen simple K-NN. Let us now look at the weighted K-NN. Let us assure we are working on a 5-NN problem. So let us assume we have to predict the class label for 'x_q'.

Let 'd_1', 'd_2', 'd_3', 'd_4' and 'd_5' be the distances of 'x_q' from the 5 nearest neighbors 'x_1', 'x_2', 'x_3', 'x_4' and 'x_5' respectively. Let 'x_1' and 'x_2' be the negative points and 'x_3', 'x_4' and 'x_5' be the positive points. Let us assume

Distance between 'x_1' and 'x_q' = d_1 = 0.1
Distance between 'x_2' and 'x_q' = d_2 = 0.2
Distance between 'x_3' and 'x_q' = d_3 = 1
Distance between 'x_4' and 'x_q' = d_4 = 2
Distance between 'x_5' and 'x_q' = d_5 = 3

So now we need to calculate the weights. Let us consider the weighted function for the a point 'x_i' as $\mathbf{w_i = 1/d_i}$

As the points 'x_1' and 'x_2' are the negative points, $w_1 + w_2 = (1/0.1) + (1/0.2) = 10 + 5 = 15$

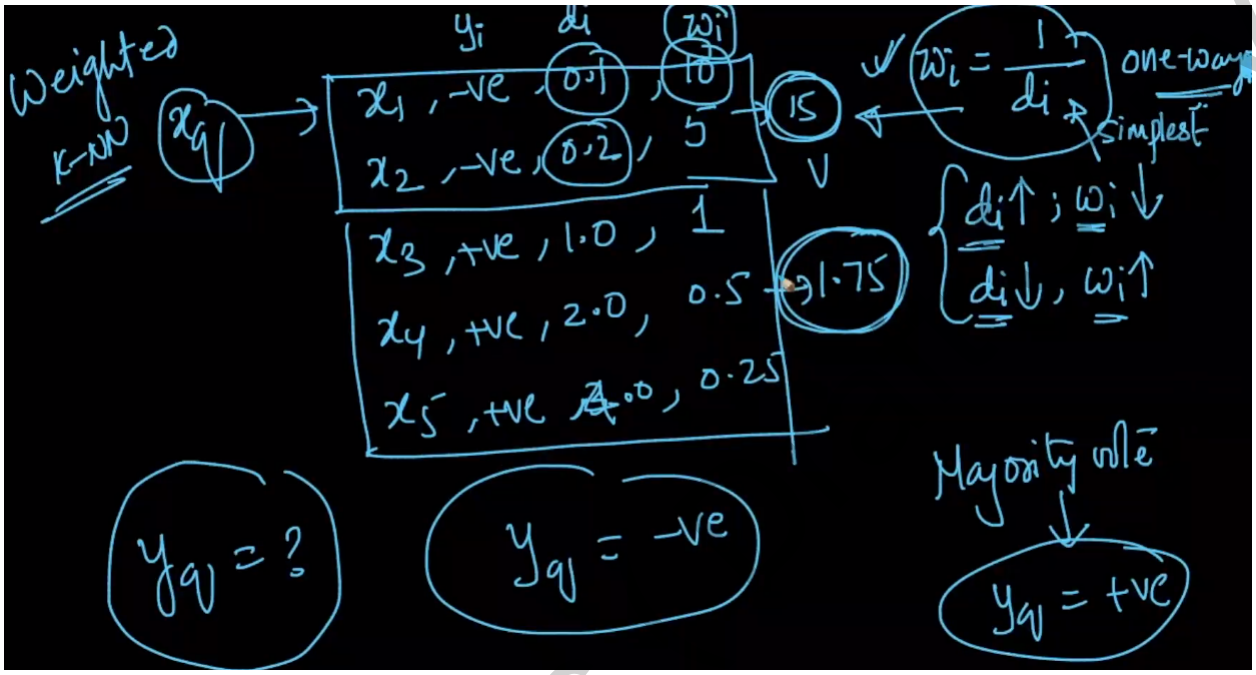
As the points 'x_3', 'x_4' and 'x_5' are the positive points, $w_1 + w_2 + w_3 = (1/1) + (1/2) + (1/3) = 1 + 0.5 + 0.33 = 1.83$

So in simple K-NN, we just go blindly with majority voting. So 'x_q' will be classified as 'positive' (as the majority count is 'positive')

In the weights K-NN, we go with the weights of the classes. So

Total Weight of the 'negative' class = $w_1 + w_2 = 15$
Total Weight of the 'positive' class = $w_3 + w_4 + w4 = 1.83$

So in the neighborhood of 'x_q', the total weight of the 'negative' class points is more than the total weight of the 'positive' class points. Hence the point 'x_q' is classified as 'negative' by weighted K-NN.



Q) When should we choose Weighted K-NN over Simple K-NN?

Ans) Weighted K-NN is used when we want to prefer closer points over farther points among the 'K' neighbors, in our decision making.

Whenever there is a requirement such that the nearest points among the nearest neighbors should contribute more in predicting the class label, then we have to go for Weighted K-NN.

The main disadvantage with Weighted K-NN is we give too much importance to the nearby points, which could result in errors in some instances. Weighted K-NN is problem specific.

Effect of Outliers on Weighted K-NN

Even the weighted K-NN gets easily affected by the presence of the outliers. Hence it is recommended to remove the outliers before applying any model.

Along with the outliers, the scale of the features also easily affects the Weighted KNN model. So we should also make sure all the features are scaled properly, before applying any model.