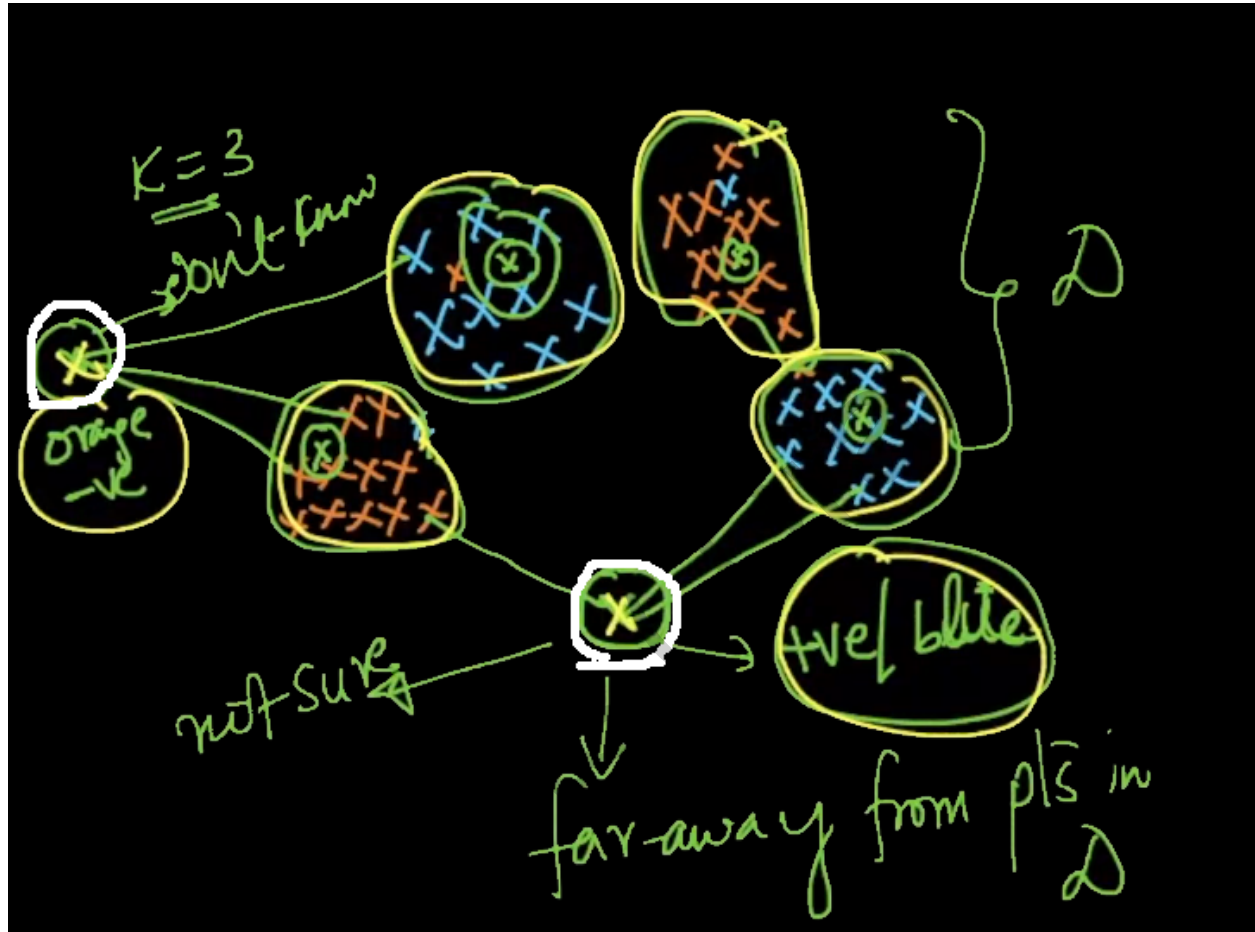


29.5 Failure Cases of K-NN

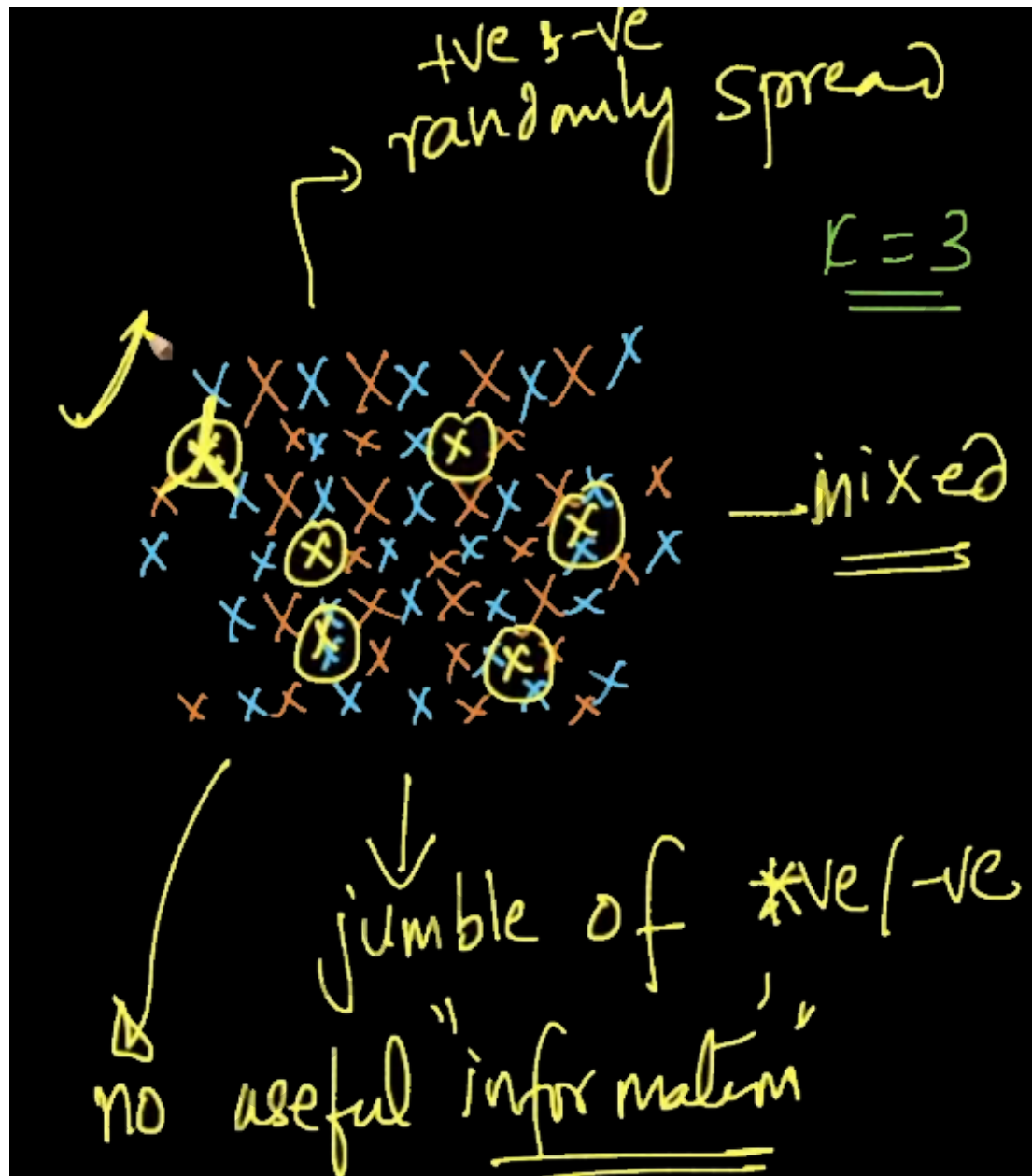
Case 1



If a point ' x_q ' is surrounded by (or) nearby more number of '+ve' points cluster or '-ve' points cluster, then we can easily classify this point into one of them, depending on the majority vote of the class labels of the points in the neighborhood.

Now what if our query points are far away from all the points in the dataset like the two given query points circled in white. When we have a query point that is far away from the rest of the points in the dataset, then we are not sure which class this point has to be classified. Though the majority of the points which are nearer belong to a particular class, we still couldn't say anything about the class accurately, because as this particular point is far away from the regular points, it might exhibit a different behaviour (ie., it doesn't belong to the pattern of the points)

Case 2



If the '+ve' and the '-ve' points are randomly spread, here we do not have exact and useful information. Here even if we predict the class label for the point 'x_q' as '+ve' (or) '-ve', it will result in an error. In this kind of situation where there is no useful information, KNN fails.

The above 2 mentioned cases are the simplest cases, where KNN fails.

Note: When the data points of different classes are randomly jumbled up, no model can work well. All models will fail. If a real world dataset shows this sort of behavior, it is better to change the feature set, as the current features are useless for classification.