

**There are 2 failure cases of k-NN:**

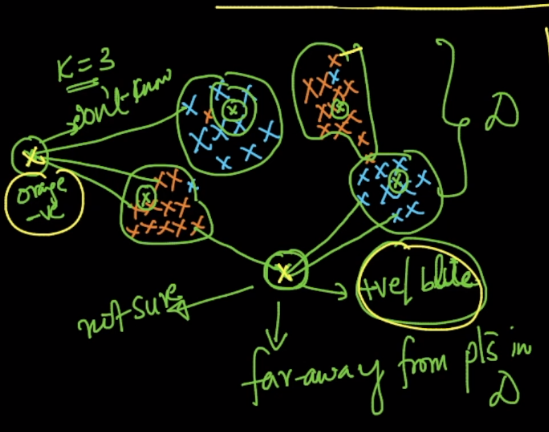
**1 ) when query point is too far from training data.**

Here as we can see that training data is useful giving information, as it clearly classifies blue and orange.

But if our query point is too far from training data, then it gives any random guess but we can’t be sure of the output and it’s better to say **don’t know** in such case.

In such cases, KNN predicts the class labels for such data points also but it may or may not be correct because  we have a pattern of points at one side and only one point on the other side that too very far away from this pattern. In this case, the K nearest neighbors are computed from this existing pattern of points only. As this individual query point is far away from the other existing points, though the class label is predicted using majority vote, it is not necessary that this point exhibits the properties of the points in the pattern/cluster. Hence we say KNN fails in such situations.

**In this kind of scenarios, algorithms like Logistic Regression, SVM, etc will work better as these algorithms just construct a hyper-plane that separates the classes and making a decision based on the geometric location of a point from the hyper-plane gives better & a confident result when compared to predicting the label from K nearest neighbors.**



**2 ) when training data is randomly spread or jumbled up:**

As we can see that given data is randomly spread, there is no information we can get from this dataset (as we are not getting the separate clusters for each class, they are completely mixed up), so determining the class of any data point in such case would give any random guess class.

Most of the ML algorithms fail for such random spread dataset, because we are not getting any information which can be used for prediction.

Some queries are as below:  
1) In real life/case, are there scenario where we deal with such dataset to make prediction from?  
2) In case such scenario exist in real life, almost all the ML classification technique will fail to make prediction out of this type of data. So, how cases like these are handled?  
3) It's also explained in video that with the geometric interpretation we get a better interpretation of data at hand that we need to deal with. So, is visualizing the dataset a necessary step to do even above any pre-processing on the dataset, to get intuition about the data?

Answers:

1,2 - Yes, for the fully jumbled up case, no ML algorithm will work. You will have to perform feature engineering to come up with a new set of features that are more discriminative.  
3 - Yes, In the journey from analysis to data-driven outcomes, data visualization plays a very important role of presenting data in a powerful and credible way. You will see its importance in solving various case studies in our course.

