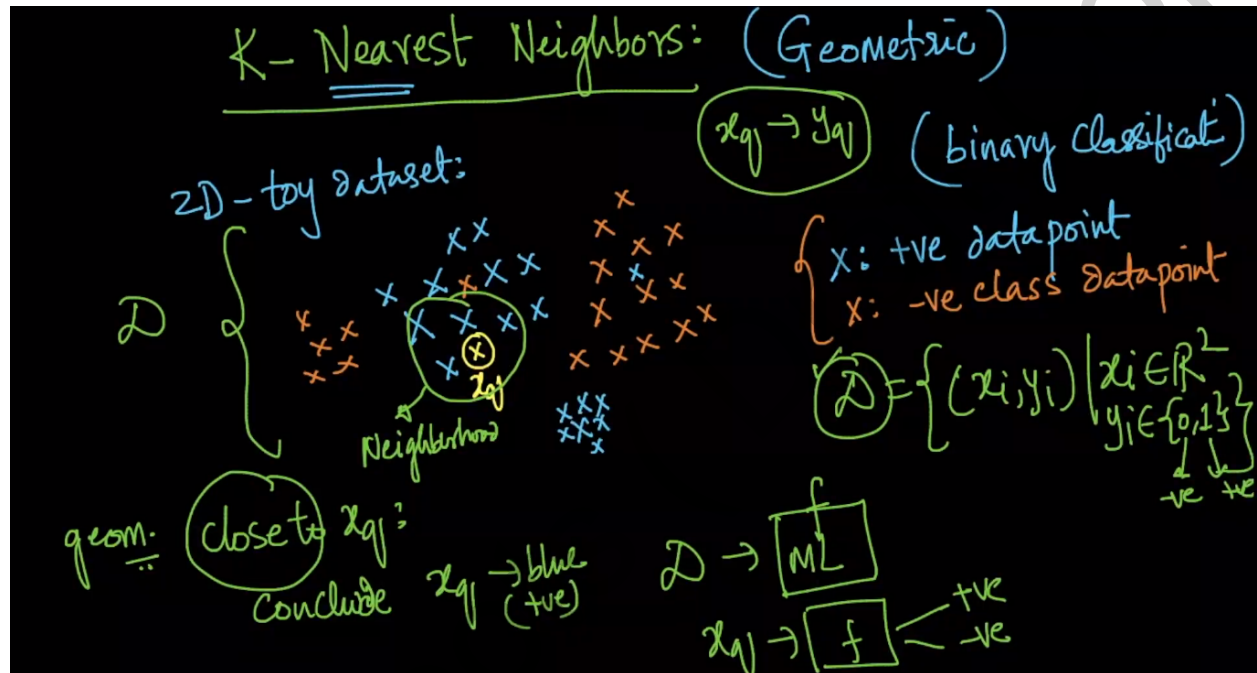


29.4 K-Nearest Neighbors Geometric Intuition with a toy example

Let us assume we are working on a binary classification problem and each observation belongs to either positive class or negative class. Let the dataset be in a 2-dimensional form. Here the dataset is represented as

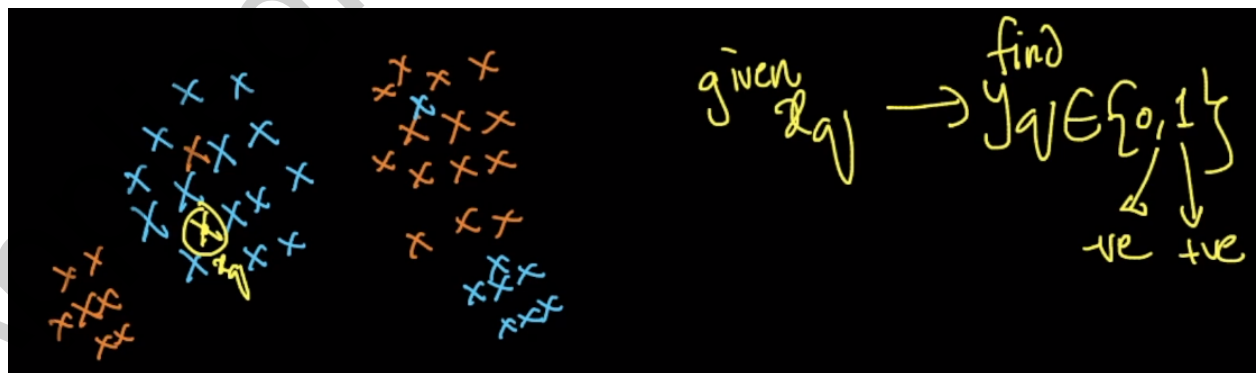
$$\mathcal{D} = \{(x_i, y_i)_{i=1}^n \mid x_i \in \mathbb{R}^2, y_i \in \{0, 1\}\}$$

The main purpose of classification is given a point ' x_q ', we have to predict the value of ' y_q '. Below is an example that was discussed starting from the timestamp 0:18.



Procedure of KNN

Given a query point ' x_q ', we have to predict the class label ' y_q '.



- 1) Compute the distance of the point ' x_q ' to all the points in the training dataset.
- 2) Sort all the distances in the ascending order, and then select the top 'K' nearest points to ' x_q '.

3) Let these 'K' points be $\{x_1, x_2, x_3, \dots, x_k\}$ and their corresponding outputs may be $\{y_1, y_2, y_3, \dots, y_k\}$.

Here for the point ' x_q ', the class to which majority of the class labels among $\{y_1, y_2, y_3, \dots, y_k\}$ belong to, will be predicted as the output for ' x_q '.

a) For example, if $K=3$ and let $\{y_1, y_2, y_3\} = \{+ve, +ve, -ve\}$, then ' x_q ' will be assigned to the '+ve' class, as the '+ve' class is in majority.

b) For example, if $K=5$ and let $\{y_1, y_2, y_3, y_4, y_5\} = \{+ve, -ve, -ve, -ve, +ve\}$, then ' x_q ' will be assigned to the '-ve' class, as the '-ve' class is in majority.

Here if 'K' is an even number, and if both the '+ve' and '-ve' class points in the neighborhood are equal in number, then it becomes difficult for the model to decide which class the datapoint ' x_q ' should be assigned to. So it is always better to avoid even values of 'K' in K-NN.

In cases where the 'K' value is even and the number of data points in '+ve' and '-ve' classes equal, then we should better increase the 'K' value, and repeat the same operation. Pick the class label with the majority of the points and assign it to ' x_q '.