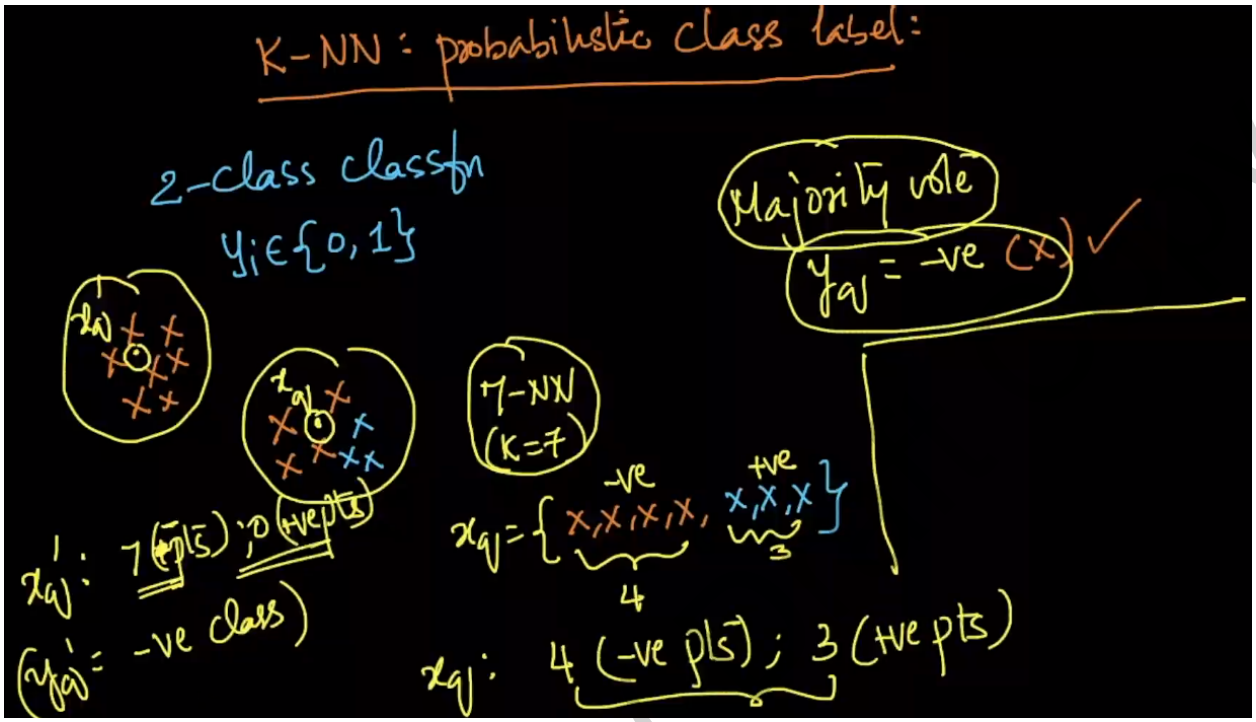
29.29 Probabilistic Class Label



Let us consider a binary classification problem. (ie., yi∈{0,1}). Let us assume we have the points as shown above, and are working on a 7-NN problem. (K=7)
So the 7 nearest neighbors of 'x_q' are

x_q = {-ve, -ve, -ve, -ve, +ve, +ve, +ve}

Here if we go with the majority vote count, we get **y_q = -ve**.

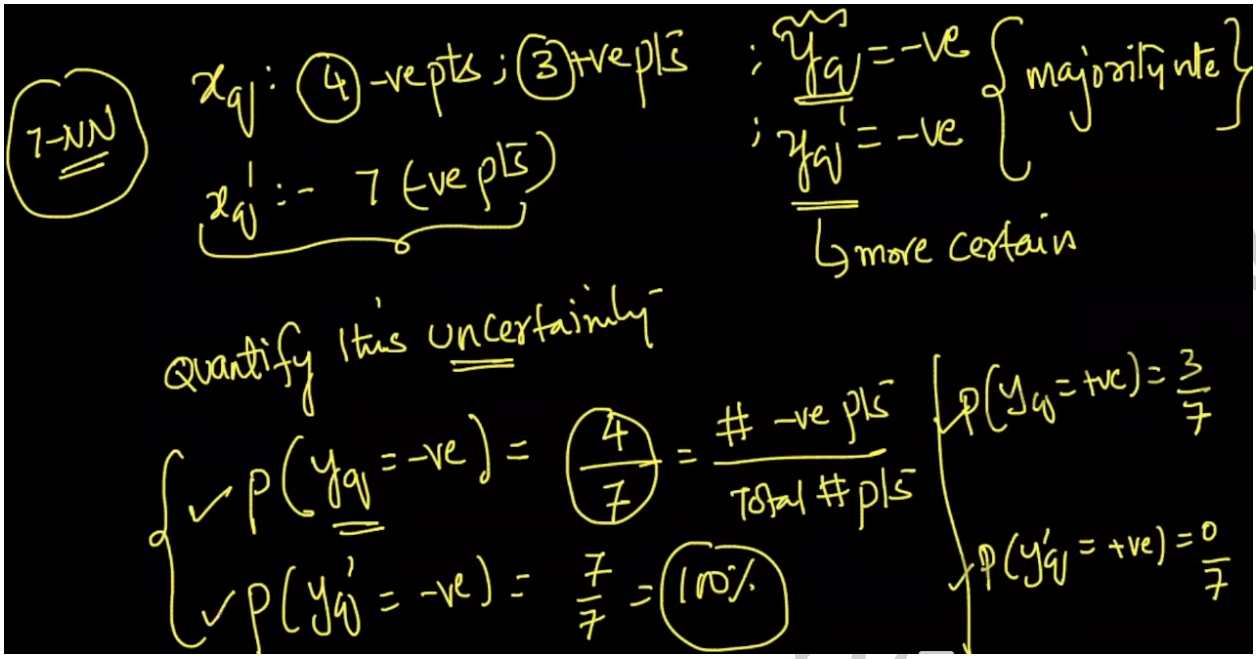
Let us assume we have another point '**x_q**'' and its 7 nearest neighbors are

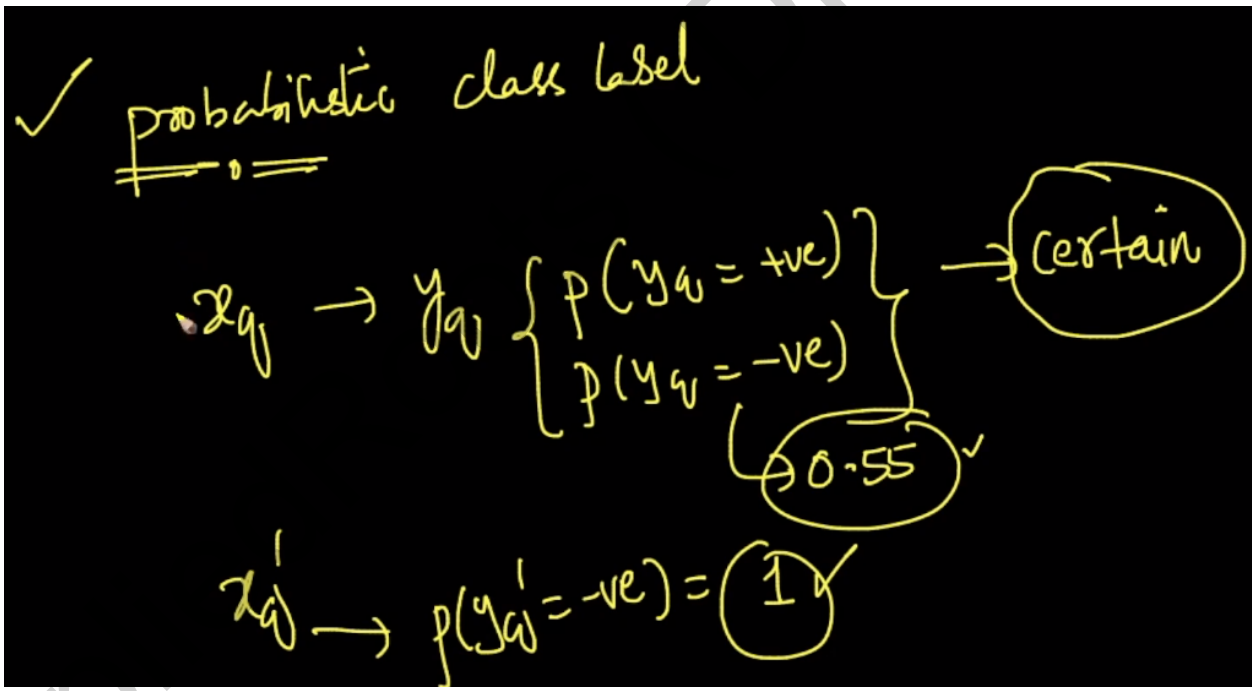
x_q' = {-ve, -ve, -ve, -ve, -ve, -ve, -ve}

Here if we go with the majority vote count, we get **y_q' = -ve**.

P(yq = -ve) = (Number of '-ve' neighbors of x_q)/(Total number of points in the neighborhood of x_q) = 4/7
P(yq = +ve) = (Number of '+ve' neighbors of x_q)/(Total number of points in the neighborhood of x_q) = 3/7

P(yq' = -ve) = (Number of '-ve' neighbors of x_q')/(Total number of points in the neighborhood of x_q') = 7/7 = 1
P(yq' = +ve) = (Number of '+ve' neighbors of x_q')/(Total number of points in the neighborhood of x_q') = 0/7 = 0





Probabilistic class labels are mostly used in classification tasks. Instead of giving deterministic class labels, we are giving probabilistic class labels. Sometimes giving probabilistic class labels is better than giving deterministic class labels.

Probabilistic result is a mathematical way to show how certain we are in predicting the class labels.

Note: We are not giving any notes for the video lectures 29.30 and 29.31 as they both are of only the code discussions.
You can download the ipython notebooks from the link given below.
https://drive.google.com/drive/folders/1tMYRWzbrSMxQ7aQ5mc8Qf4190gPSt5fy

For any queries, please feel free to post them in the comments section below the video lecture.