KNN, K-nearest neighbours.

Estimates conditional distribution of y given x, and then classifies a given observation to the class with highest probability.

Step 1: Pick integer K.

Step 2: x_0 (test observation)

identifying K points in the data that are closet to the x_0. (N_0)

Step 3: it estimates conditional probability for each class j as the fraction of points in N_0, whose response values equals j.

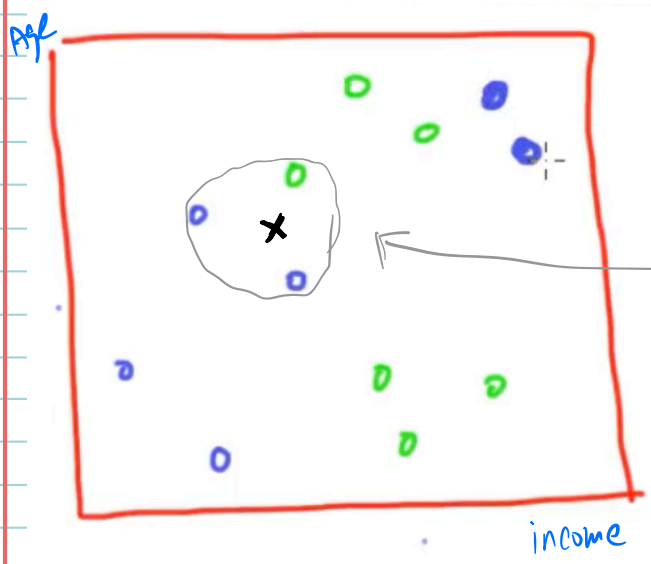
$$Pr(y=j|x=x_0) = \frac{1}{K}\sum_{i \in N_0} I(y_i=j).$$

proportion of each class in the neibourhood of x_0.

for every point in the neibourhood

Step 4: Applyes Bayer's rule and classifiers x_0 to the

class with largest probability in the Neibourhood.

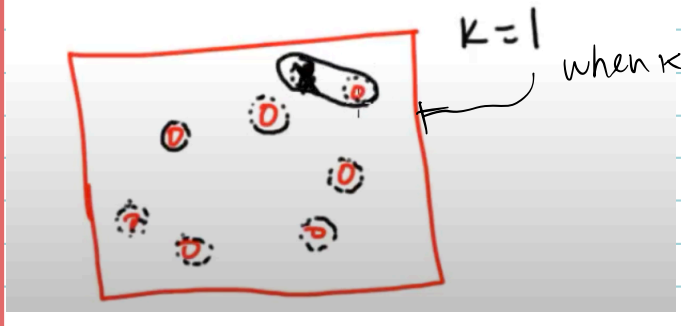


① let $k=3$. choose 3 points closet to x_0.

② 1 green
2 blue

③ $P(\text{green}) = \frac{1}{3}$.
$P(\text{blue}) = \frac{2}{3}$.

④ we classify x as blue
cause $P_{blue} > P_{green}$



when $k=1$. So many errors : overfitting.

$k=1$ overfitting

$k=7$ useless → includes all the data points, so useless.