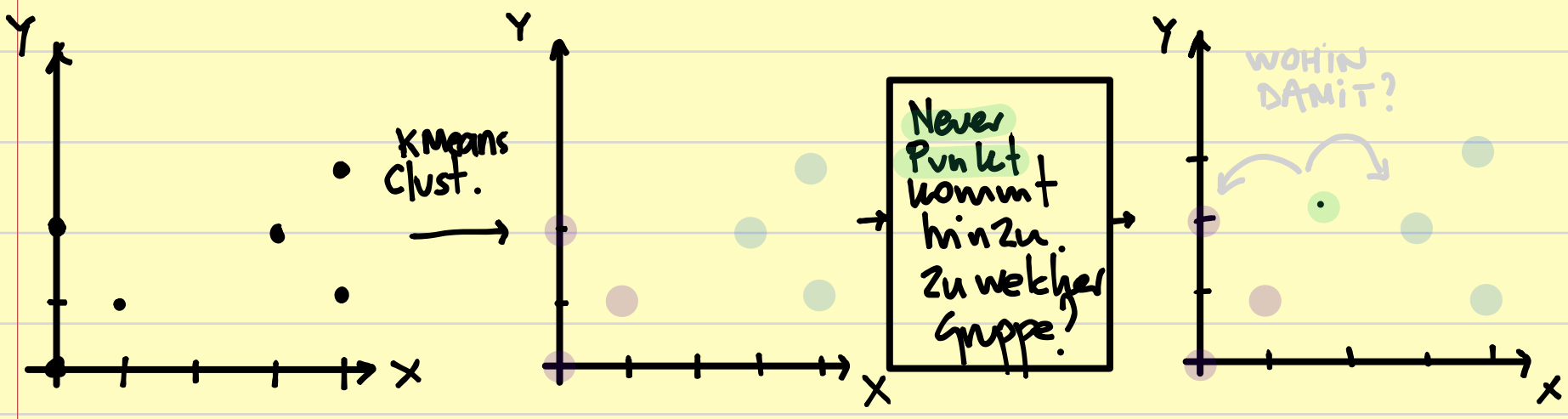
K-Nearest Neighbour (KNN)



X	Y
0	0
0	2
1	1
3	2
4	1
4	3

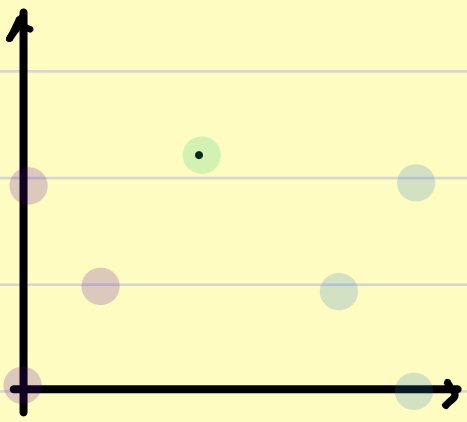
Introduction

KNN is a type of supervised learning algorythm used for both regression & classification.

KNN tries to predict the correct class for the test data. Then select the K. number of points which are closest to the test data. The KNN algorithm calculates the probability of the test data belonging to the classes, and assigns the class to the highest probability. In the case of regression, the value is the mean of the „K" selected training points.

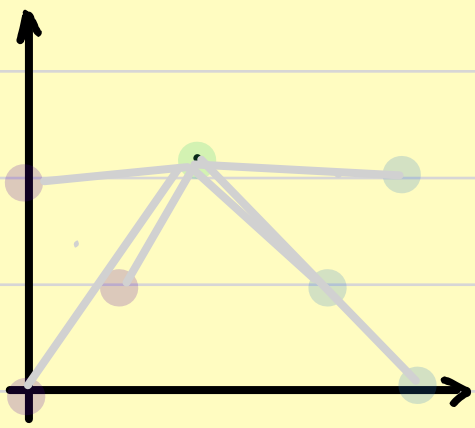
Why do we need a KNN algorythm? Suppose we have two catogories, and we have a new data point, so this data point will lie in one of the categories. To decide in which category it belongs, we use KNN.

0. Look at the data



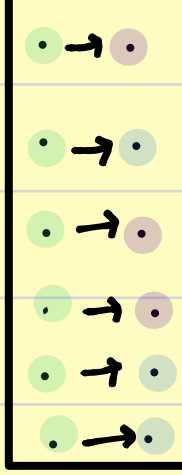
We want to classify the new data point.

1. Calculate distances



Start by calculating the distances between the green point and all other points.

2. Find the neighbors



→ 1.NN

→ 2.NN

Next, find the nearest neighbors by ranking points with increasing distance.

NN: Nearest Neigbor

3. K

We decide „K" as the number of neighbors

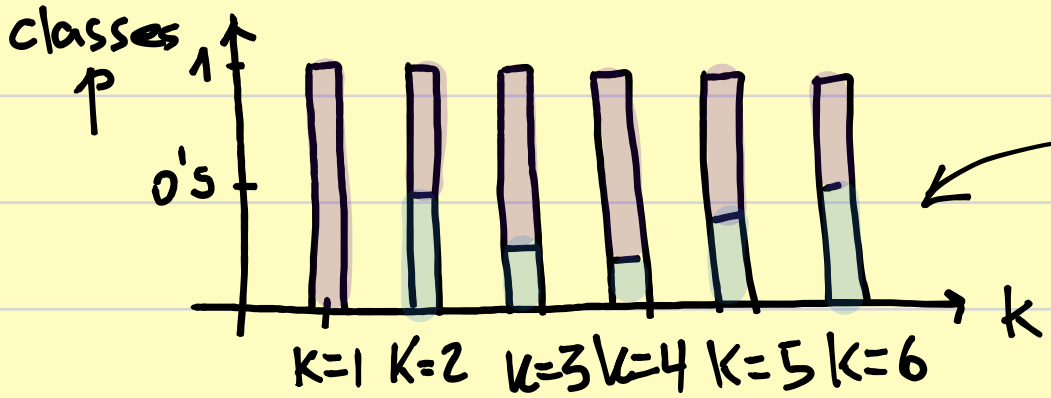
4. Vote the labels

k=4

→ 3 votes

→ 1 vote

Vote on the predicted class labels based on the classes of the k nearest neighbour.



This plot shows that the minimum error is found on K = 4.

How do we select the optimal k-value?

There are no predefined methods to find the most favorable value of k.

Initialize a random value of K and start computing.

Choosing a small value of K leads to unstable decission boundaries.

The substantial K value is better for classification as it leads to smoothening the decission boundaries.

Beispiel.

	X	Y
	0	0
A	0	2
	1	1
	3	2
B	4	1
	4	3

$x_1 \equiv$ Neuer Punkt $\equiv [2\,5]$

$d_{x_1,A_1} = \sqrt{(2-0)^2+(5-0)^2} = 5{,}38$ — kNN 6.

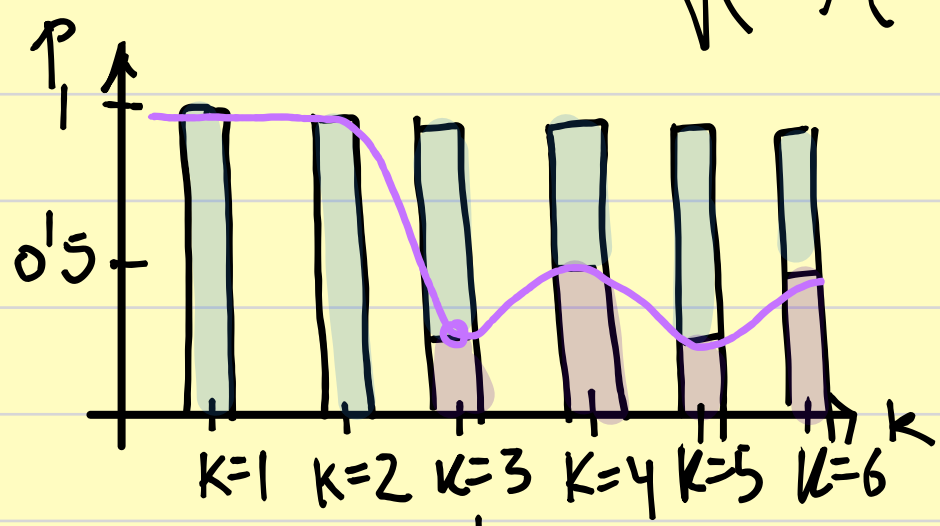
$d_{x_1,A_2} = \sqrt{(2-0)^2+(5-2)^2} = 3{,}605$ — 3.

$d_{x_1,A_3} = \sqrt{(1-2)^2+(5-1)^2} = 4{,}123$ — 4.

$d_{x_1,B_1} = \sqrt{(3-2)^2+(5-2)^2} = 3{,}16$ — 2.

$d_{x_1,B_2} = \sqrt{(4-2)^2+(1-5)^2} = 4{,}47$ — 5.

$d_{x_1,B_3} = \sqrt{(4-2)^2+(5-3)^2} = 2{,}82$ — 1.



K = 3 → Der Punkt gehört zu der ● -Gruppe.

- K-DIMENSIONAL TREE (kd Tree)
- BALL-TREE