

KNN (K nearest Neighbors) Classification:

STEP 1: Choose the number K of neighbors



STEP 2: Take the K nearest neighbors of the new data point, according to the Euclidean distance



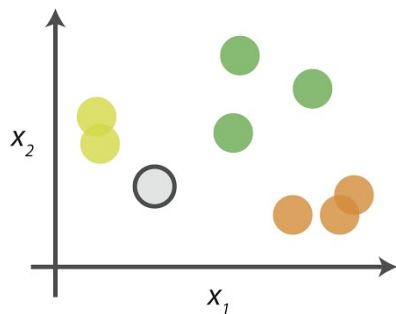
STEP 3: Among these K neighbors, count the number of data points in each category



STEP 4: Assign the new data point to the category where you counted the most neighbors

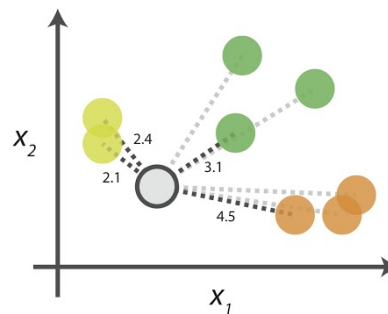
kNN Algorithm

0. Look at the data











Say you want to classify the grey point into a class. Here, there are three potential classes - lime green, green and orange.

1. Calculate distances









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

2. Find neighbours

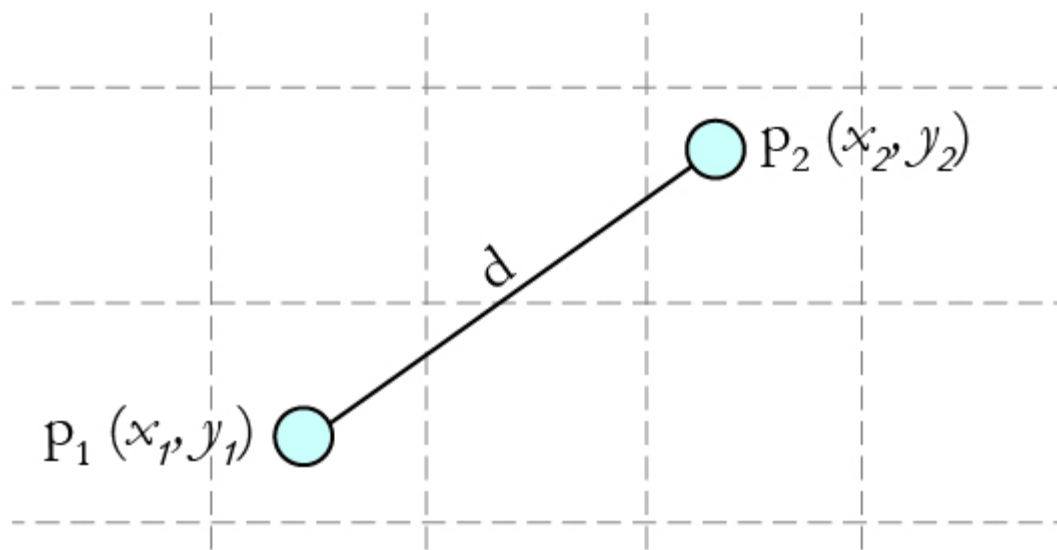
	Point	Distance	
	 	2.1	→ 1st NN
	 	2.4	→ 2nd NN
	 	3.1	→ 3rd NN
	 	4.5	→ 4th NN

Next, find the nearest neighbours by ranking points by increasing distance. The nearest neighbours (NNs) of the grey point are the ones closest in dataspace.

3. Vote on labels

Class	# of votes	
	2	➔ Class  wins the vote! Point  is therefore predicted to be of class  .
	1	
	1	

Vote on the predicted class labels based on the classes of the k nearest neighbours. Here, the labels were predicted based on the $k=3$ nearest neighbours.



$$\text{Euclidean distance (d)} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

Here

x_2, y_2 → target data points

x_1, y_1 → actual (given data point)

diameter	weight	Fruit Name
6	100	Apple
6.1	95	Apple
3	30	Banana
3.2	35	Banana
5.5	80	?

Here Target data points:

(5.5, 80)

$$D1 = \sqrt{(\sqrt{5.5-6}) + \sqrt{80-100}} = 2.5 + 400 = 402.5 = 20.06$$

$$D2 = \sqrt{(\sqrt{5.5-6.1}) + \sqrt{80-95}} = 1.84 + 225 = 226.84 = 15.06$$

$$D3 = \sqrt{(\sqrt{5.5-3}) + \sqrt{80-30}} = 6.25 + 2500 = 2506.25 = 50.06$$

$$D4 = \sqrt{(\sqrt{5.5-3.2}) + \sqrt{80-35}} = 50.29 + 2025 = 2075.29 = 45.55$$

If $k=3$,

Choose 3 nearest distance:

$D1, D2, D4$

Now see classes of those data points:

$D1 = \text{Apple}$

$D2 = \text{Apple}$

$D4 = \text{Banana}$

Voting \rightarrow Apple wins

So target datapoints will belong to Apple

Note: if we take even value for k , then there may be chance of tie if so model randomly pick class for target data point at each iteration.