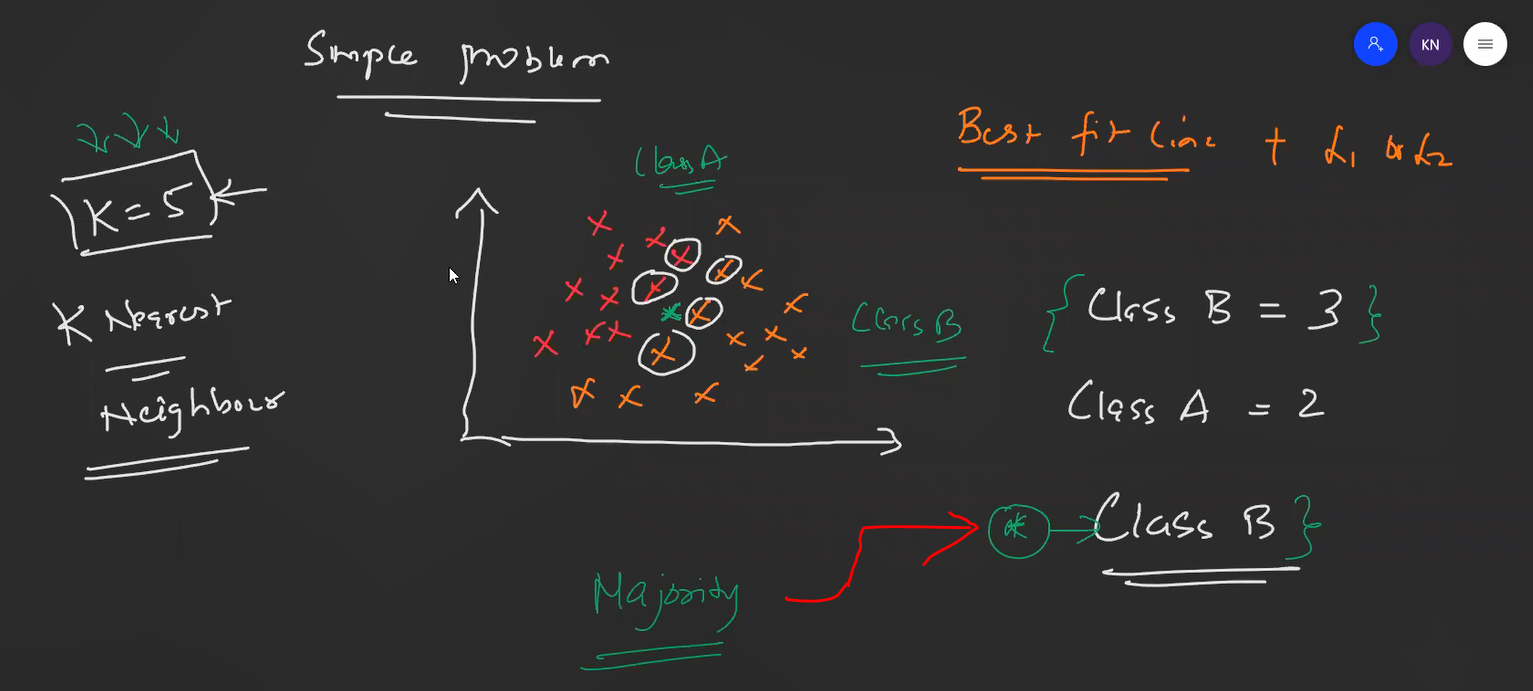
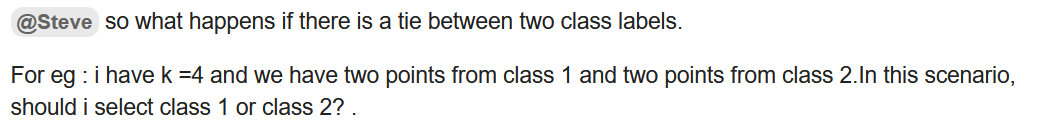
**Classification perspective**

Red points 🡪 class A  
Orange points 🡪 class B  
Want to find where the green point belongs to????

That green point belongs to Class-B since it was the majority.

**Regression perspective**

**What will be the hyper parameter ??**K Nearest Neighbour 🡪 K (See practical implementation)

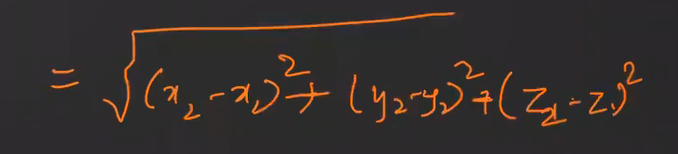
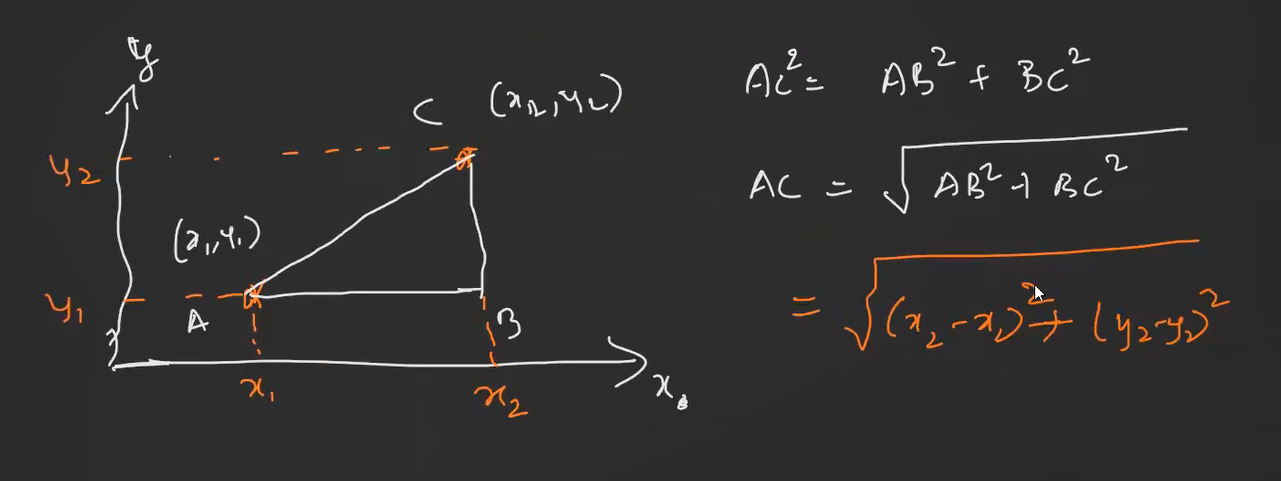
**What will be K value 🡪 Odd/Even ??**Definitely odd value, since  


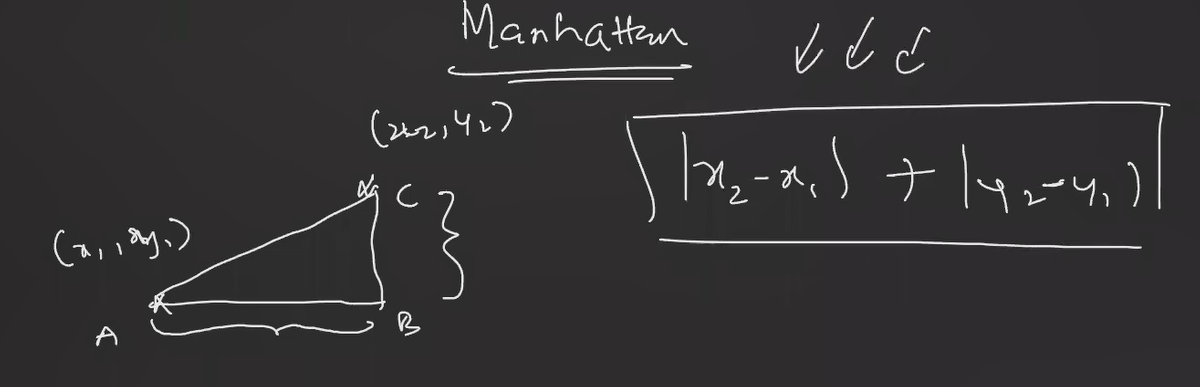
**Why KNN is not used for outliers/imbalanced data-sets ??**

Class A 🡪 white points  
Class B 🡪 orange points  
Where the new red point belongs to ?? 🡪 Its not possible to say.

With Box plot and IQR, outliers will be removed.

**Calculating Distance**

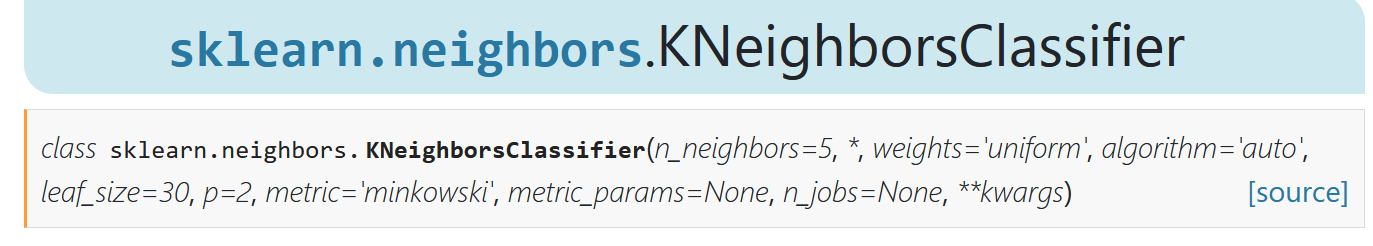
Euclidian distance from Pythagoras theorem **3D perspective**

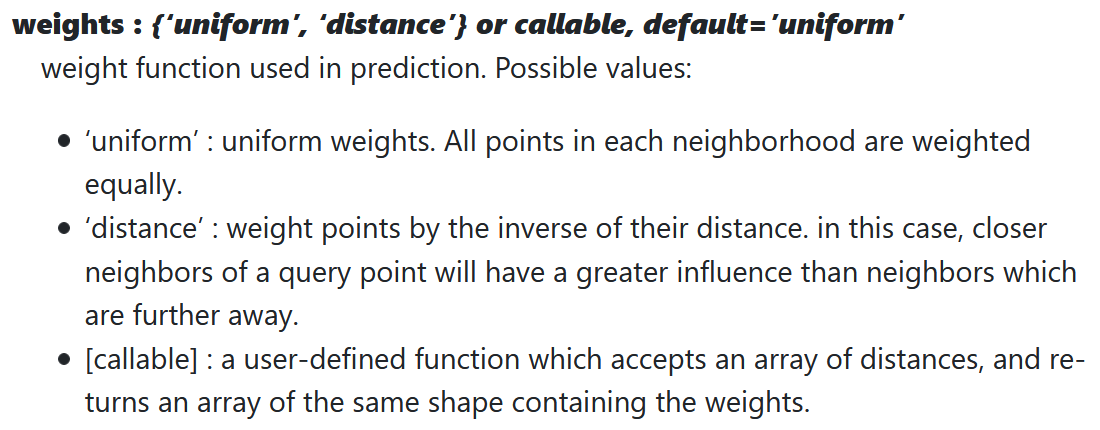
Manhattan Distance

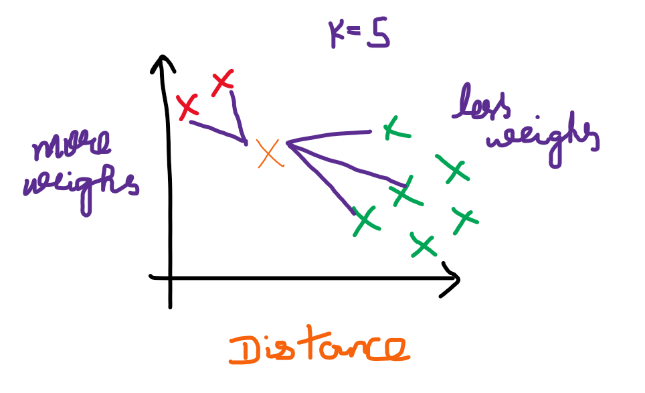
Which ever-algorithms standardization required ????  
Algorithms involving distance,

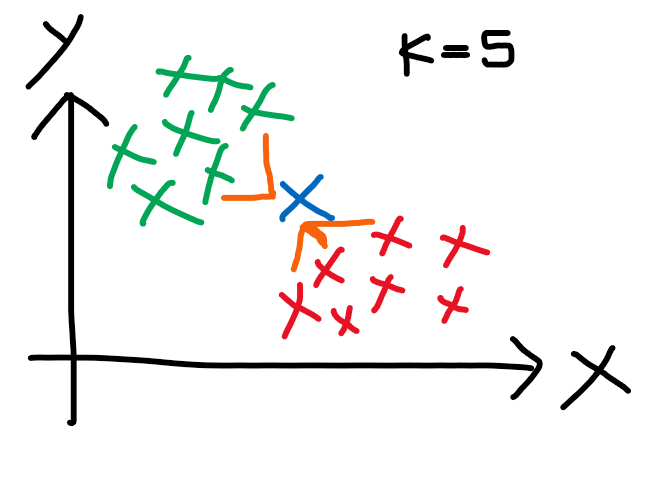
Interview Question

**KNN Scikit Learn**

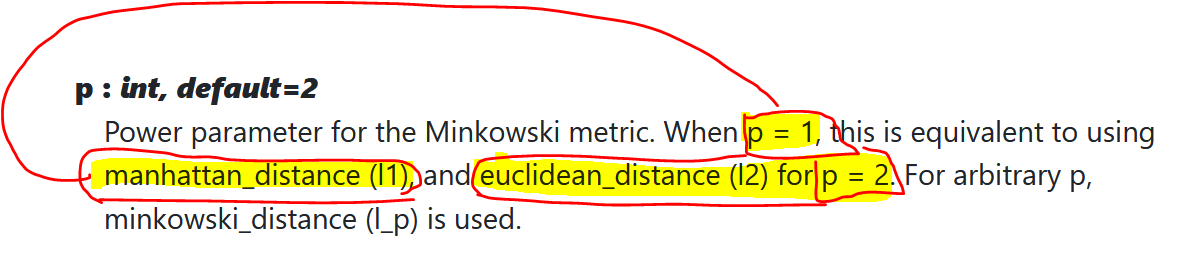




  
Distance --> for imbalanced data-set

  
Uniform 🡪 Used for Balanced Data-Set

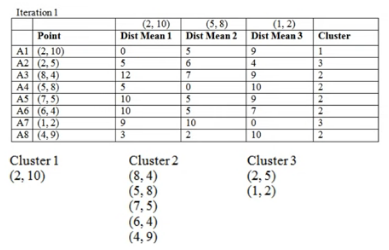
far-away points --> less weighs  
near points --> more weighs  
Take the nearest points and see the majority and assign that query point to that majority class.



# **Type-1**

# **Type-2**

## **Example-1**



Text

Description automatically generated with medium confidence

Iteration-2Table

Description automatically generated

Iteration-3  
Table

Description automatically generated

Iteration-4  
Table

Description automatically generated

Both iteration-3 and iteration-4 makes the same clusters 🡪 SO Stop  
If the centroids are matching 🡪 Don’t do iteration-4  
But (3, 9.5) and (3.67, 9) 🡪 Not equal, so do iteration-4

Calendar

Description automatically generated

Chart

Description automatically generated