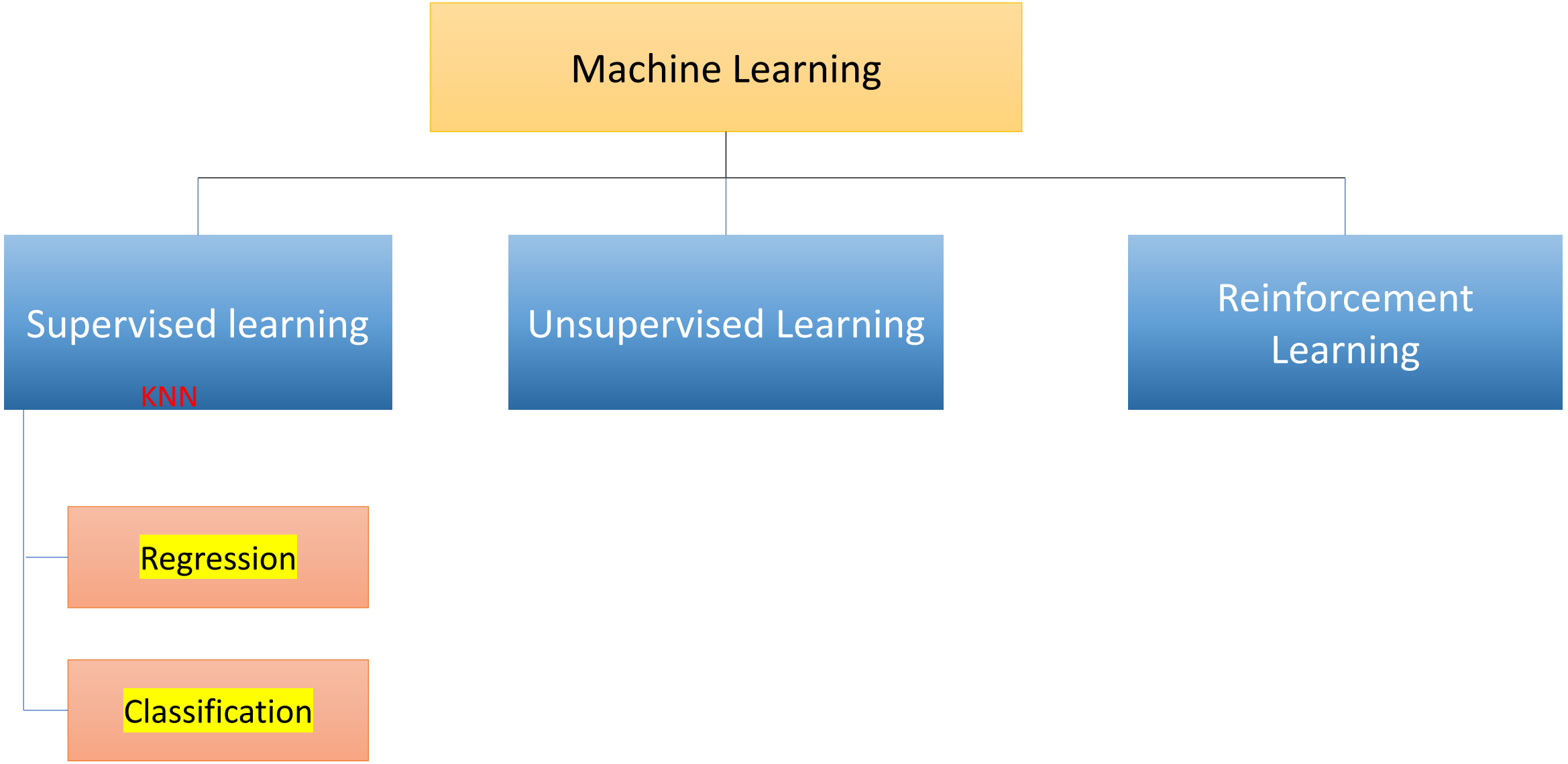
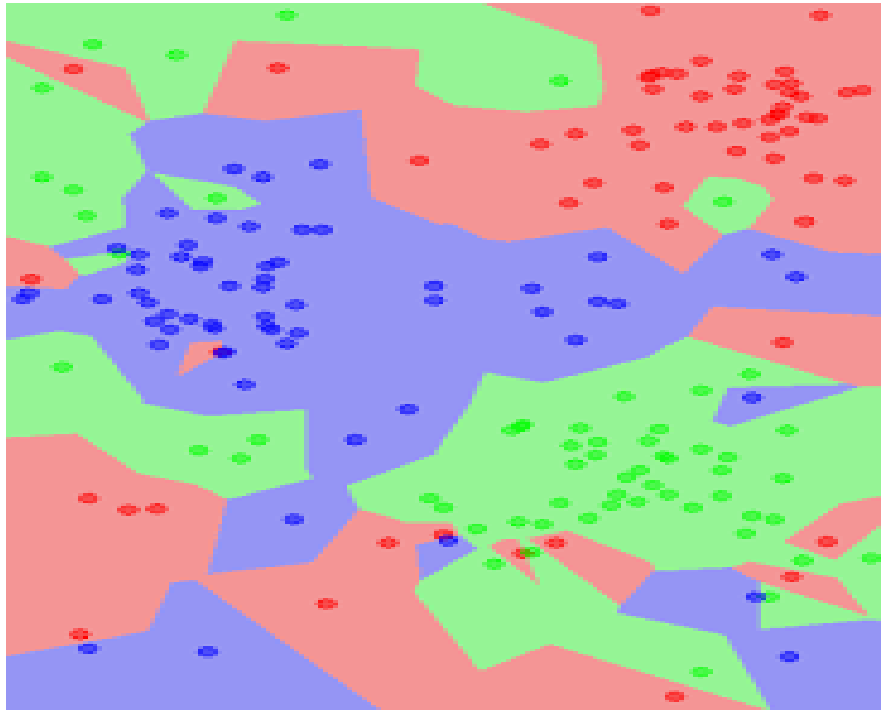
Machine Learning Type



K-Nearest Neighbor algorithm

- The KNN algorithm assumes that similar things exist in proximity. In other words, similar.
- things are near to each other.

K-Nearest Neighbor algorithm

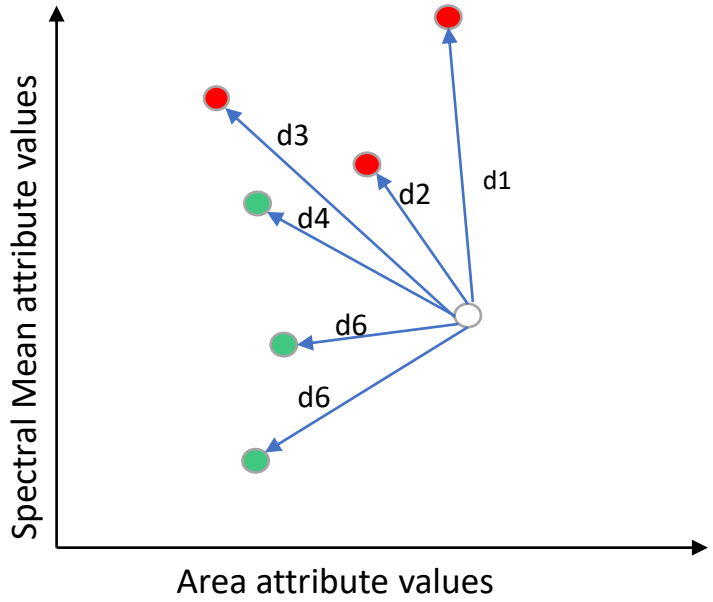


Continue

- Notice in the image that most of the time, similar data points are close to each other.
- KNN captures the idea of similarity (sometimes called distance, proximity, or closeness) with some mathematics we might have learned in our childhood- calculating the distance between points on a graph.

K-Nearest Neighbor algorithm

- One way might be preferable depending on the problem we are solving.
- However, the straight-line distance (also called the Euclidean distance)is a popular and familiar choice.



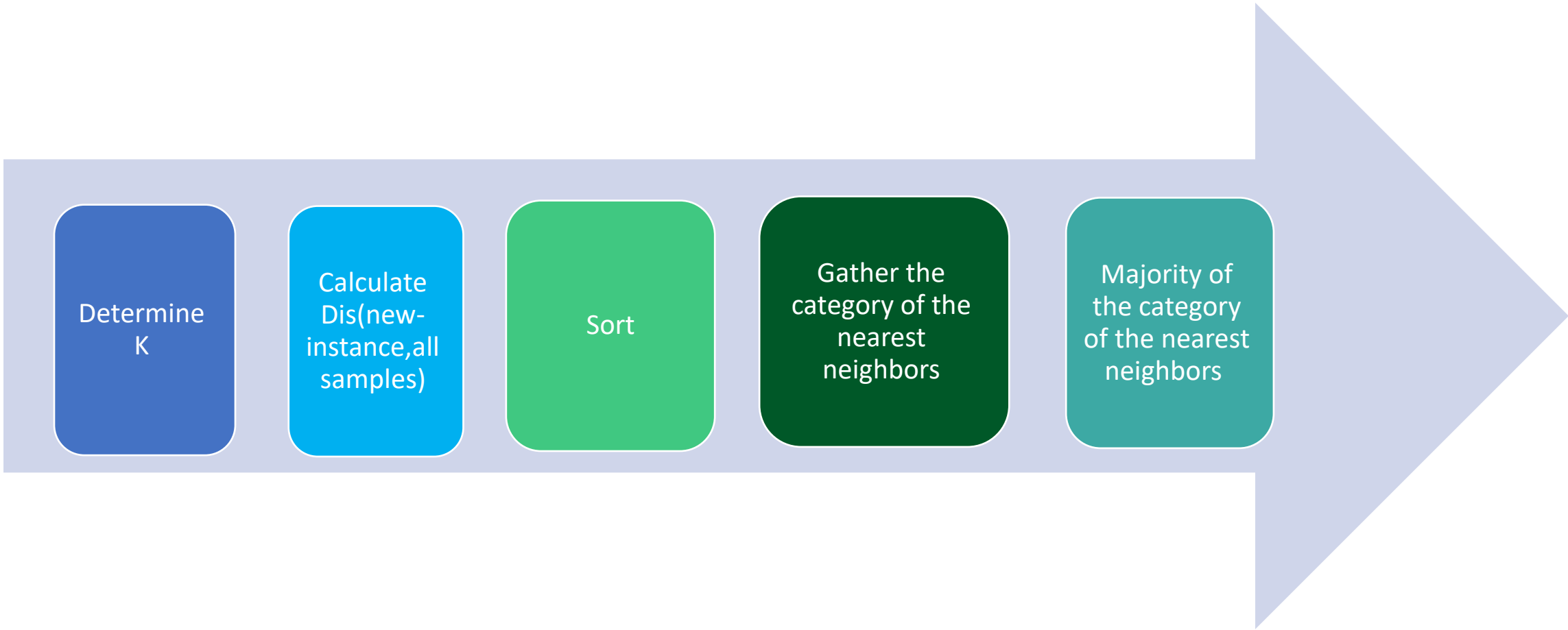
K-Nearest Neighbor algorithm

- Euclidean distance between two points:
- X=(X1,X2...,Xn)
- Y=(Y1,Y2…,Yn) is:

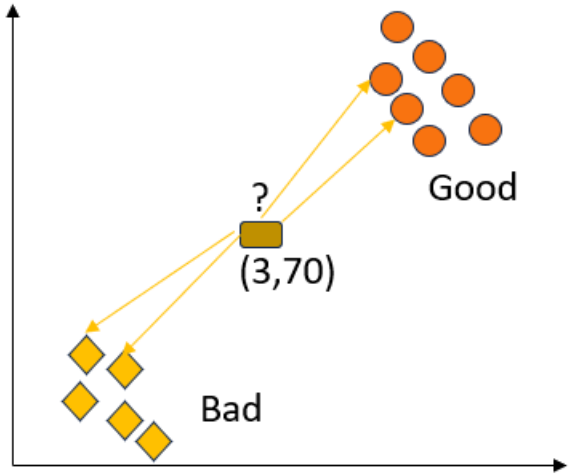
$$d(X,Y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

K-Nearest Neighbor algorithm

- Algorithm step by step:



Classification Using KNN



Name	Cigarettes	Weight	Heart Attack	Distance	
A	7	70	Bad	$\sqrt[2]{(3-7)^2+(70-70)^2}=4$	1
B	7	40	Bad	$\sqrt[2]{(3-3)^2+(7-7)^2}=30.27$	4
C	3	40	Good	$\sqrt[2]{(7-3)^2+(7-7)^2}=30.00$	2
D	1	40	Good	$\sqrt[2]{(7-3)^2+(7-7)^2}=30.07$	3
E	3	70	Bad		

K-Nearest Neighbor algorithm

-KNN Features:

1. Simplicity of implementation: KNN is a simple and easy algorithm to implement. It does not require expensive training, as only all training data is stored.

2. KNN does not make specific assumptions about the data, but rather relies on directly available data.

3. Ability to handle non-linear data: KNN can handle non-linear data which may be difficult to classify by other algorithms

K-Nearest Neighbor algorithm

- Disadvantages of KNN:

1. The presence of single values or a false value in the data can have a significant impact on the performance of KNN. Single values can cause misclassification of points.

2. Storage cost: All training data must be stored in memory to be used for prediction. This means that in the case of large sets of data, it can consume a lot of storage space.