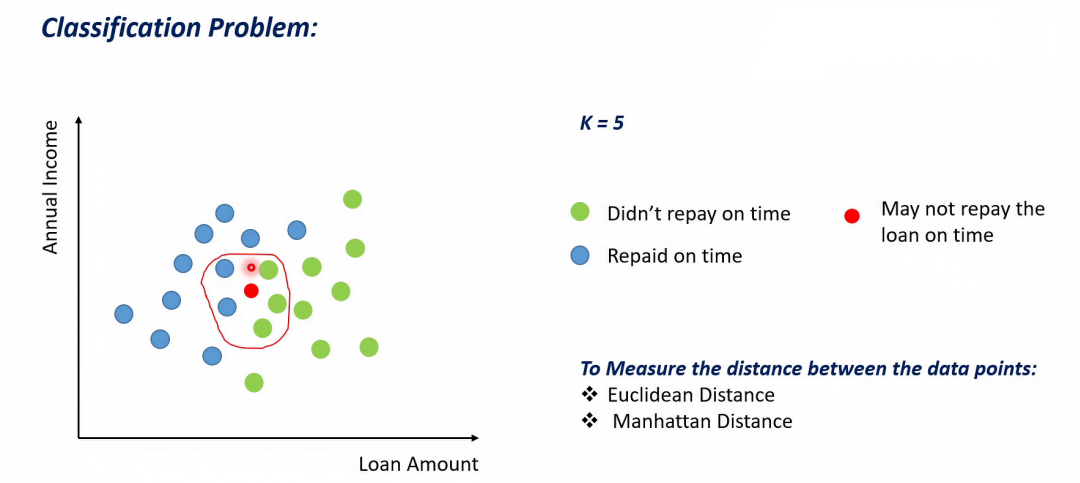
K- Nearest Neighbors

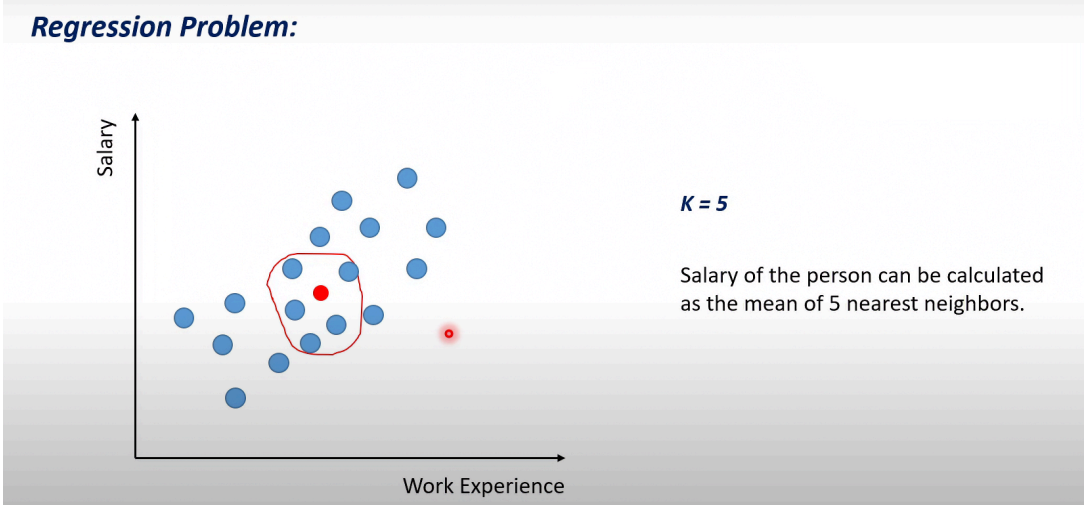
Created	@August 8, 2025
Edited	@August 8, 2025 8:07 AM
Archive	☐

1. It is a supervised learning model
2. Used for both classification and regression
3. Can be used for non-linear data
4. We have to define the K-Value

For classification:



For Regression:



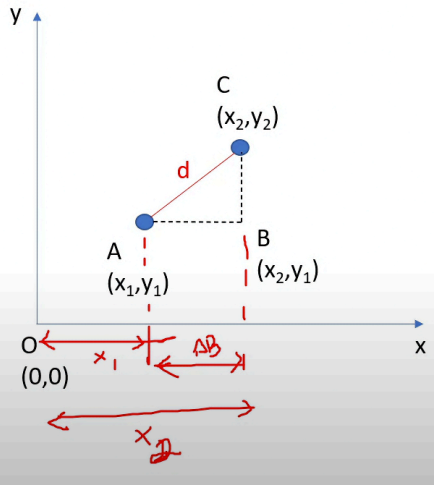
Advantages:

1. Works well with smaller datasets with less number of features.
2. Can be used for both classification and regression
3. Easy to implement for multi class classification problems
4. Different distance criteria can be used (Euclidean distance, Manhattan distance etc.)

Disadvantages:

1. Choosing the optimum K value
2. Less efficient with high dimensional data
3. Does not perform well on imbalanced dataset
4. Sensitive to Outliers.

Euclidean Distance:



Pythagoras Theorem:

$$AC^2 = AB^2 + BC^2$$

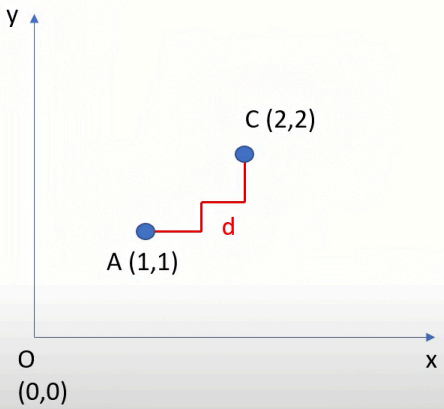
$$AC = \sqrt{AB^2 + BC^2}$$

$$AC = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

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

This distance "d" is called the Euclidean Distance.

Manhattan Distance:



Manhattan Distance formula:

$$d = |x_1 - x_2| + |y_1 - y_2|$$

(x_1,y_1) = A (1,1)

(x_2,y_2) = B (2,2)

$$d = |1 - 2| + |1 - 2|$$

$$d = 1 + 1$$

$$d = 2$$

Manhattan distance is preferred over Euclidean distance when there is high dimensionality in the data.