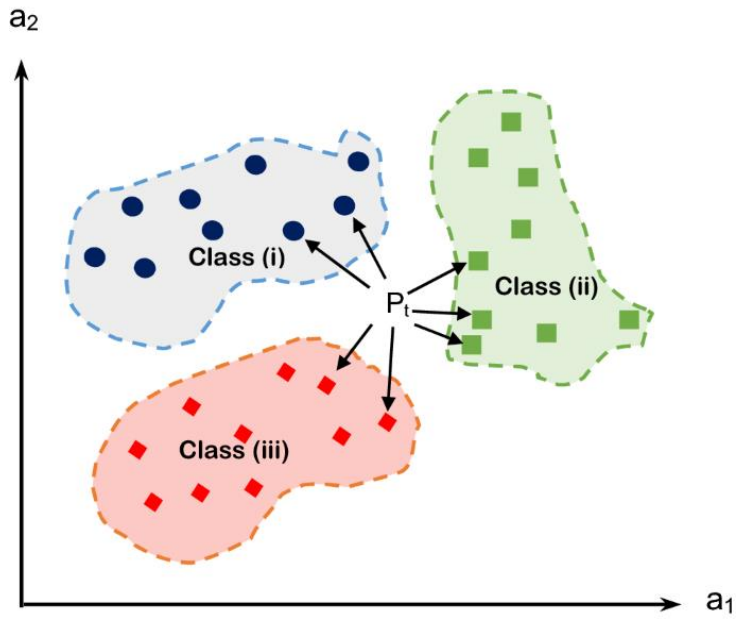
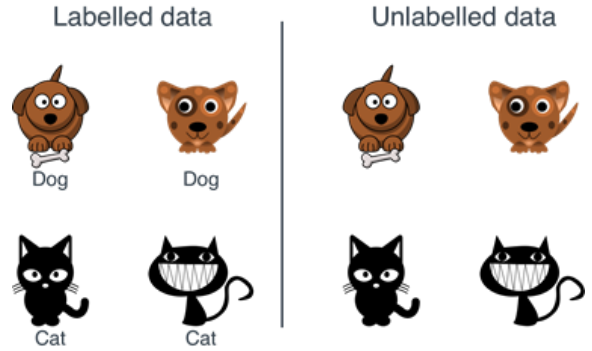
Siddhardhan

# Math behind K-Nearest Neighbors (KNN) Classifier

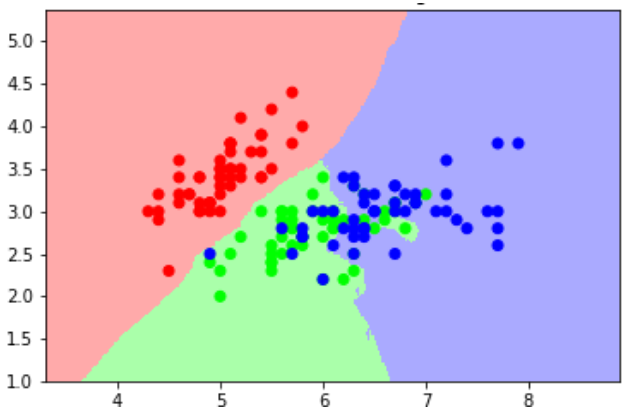


# *K-Nearest Neighbors*

## *About K-Nearest Neighbors:*

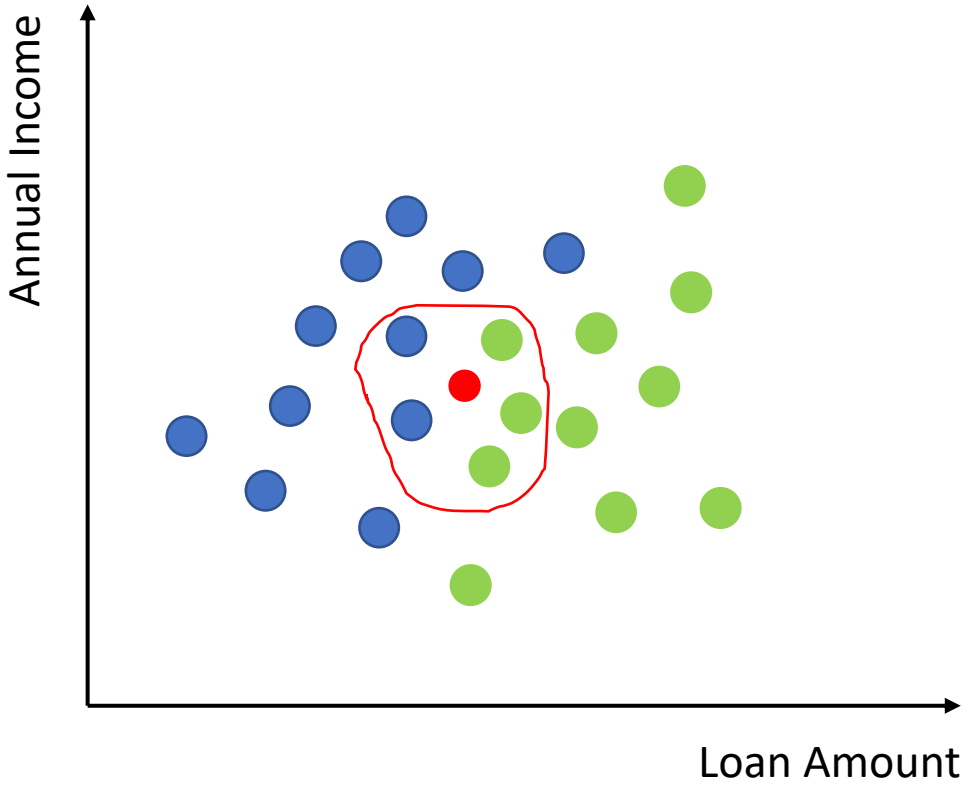



1. Supervised Learning Model
2. Used for both Classification & Regression
3. Can be used for non-linear data
4. K - Neighbors



# K-Nearest Neighbors

## Classification Problem:



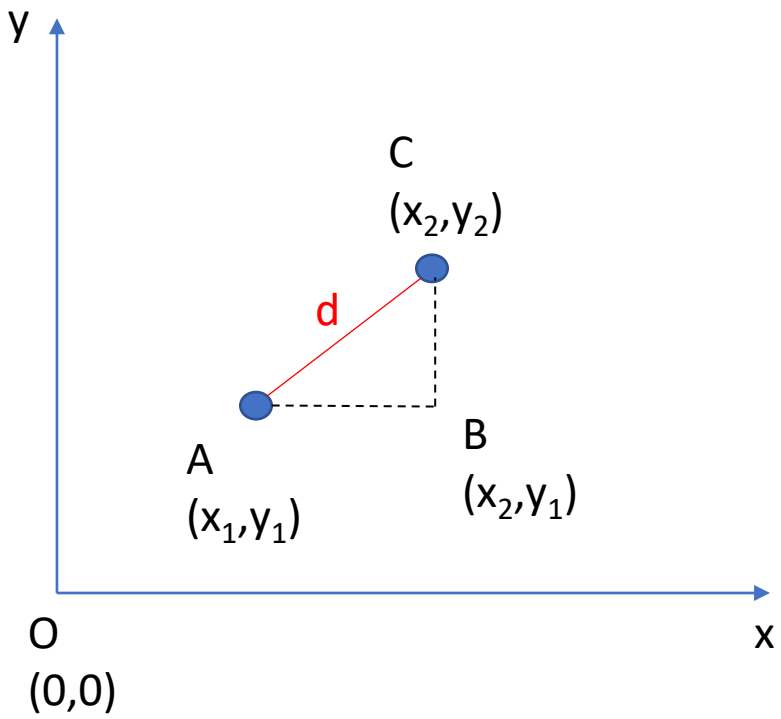

***K = 5***

- Didn't repay on time
- Repaid on time
- May not repay the loan on time

***To Measure the distance between the data points:***

- Euclidean Distance
- Manhattan Distance

# *Euclidean Distance*




**Pythagoras Theorem:**

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

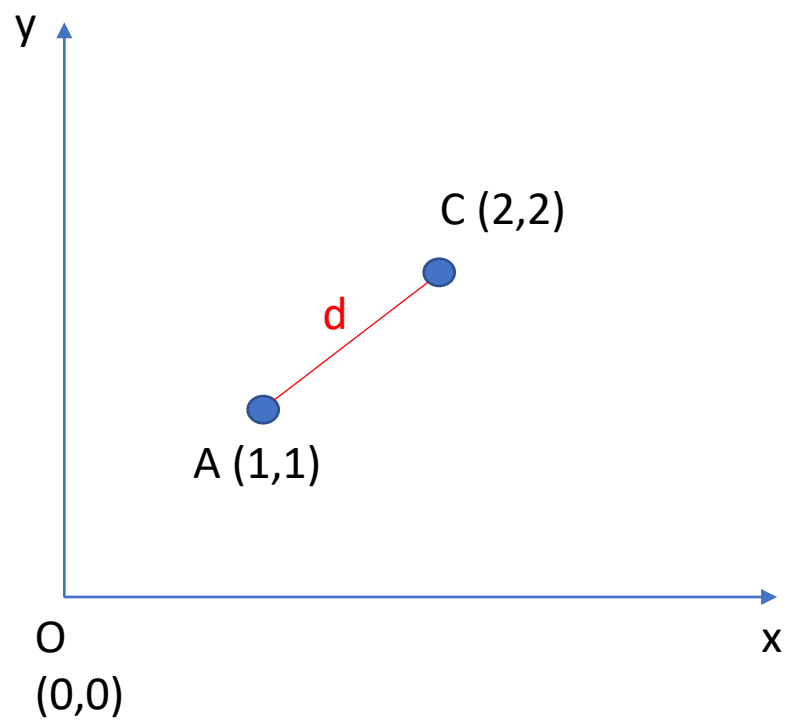
$$AC = \sqrt{AB^2 + BC^2}$$

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

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

This distance “d” is called the Euclidean Distance.

# *Euclidean Distance*




*Euclidean Distance formula:*

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

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

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

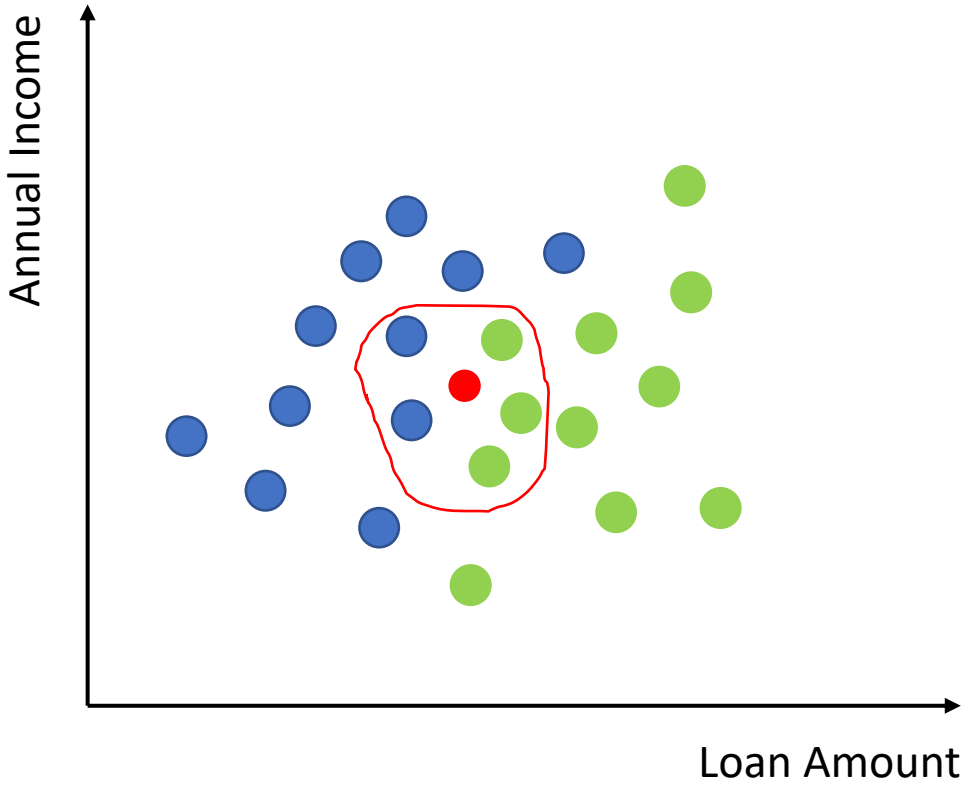
$$d = \sqrt{(2 - 1)^2 + (2 - 1)^2}$$

$$d = \sqrt{1 + 1}$$

$$d = \sqrt{2}$$

# K-Nearest Neighbors

## Classification Problem:



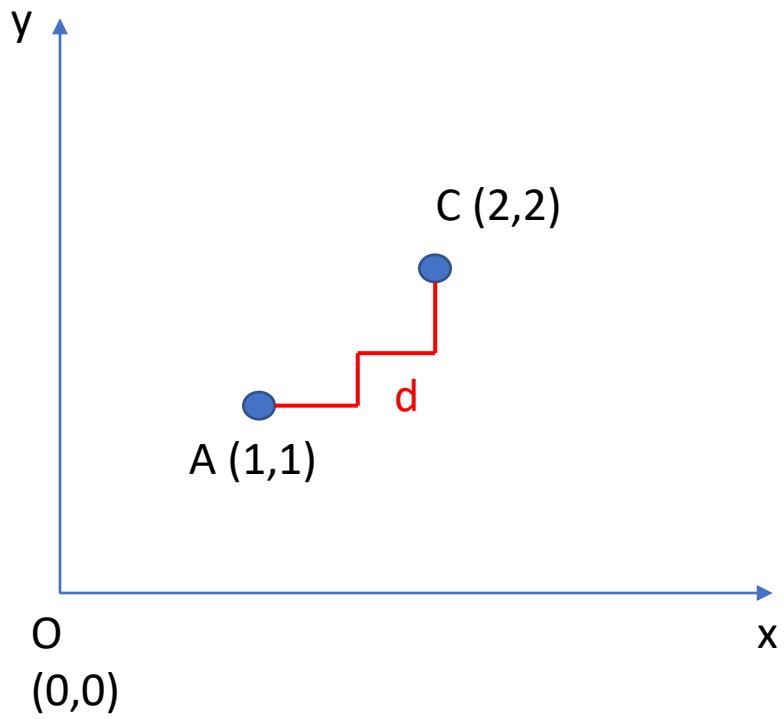

***K = 5***

- Didn't repay on time
- Repaid on time
- May not repay the loan on time

***To Measure the distance between the data points:***

- Euclidean Distance
- Manhattan 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.