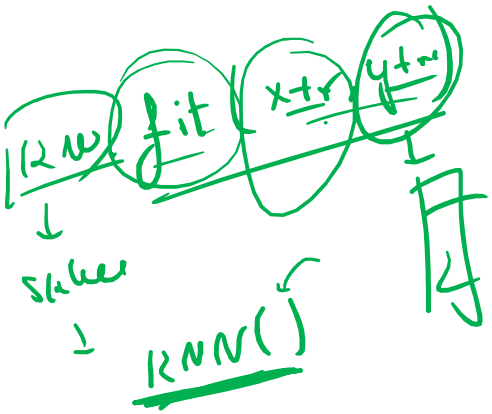
→ K-NN (K - Nearest Neighbours)

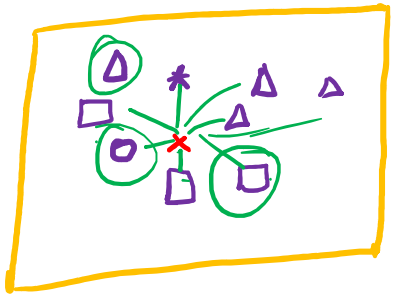
↓

Arbitrary value

→ Classification → Supervised ML Model (Feature, Target)

→ Concept of Neighbours → Nearby data points





→ Does Not have a training process

Only prediction process is there

↓↓

lazy - learning

K → [1, N]

K = 1 → Pred = □

K = 2 → Pred = X Tie 1 - □, 1 - □

K = 3 → Pred = X 1 - □, 2 - □, 1 - △

K = 4 → Pred = □

K = 5 → Pred = □

K = 6 → Pred = X

K = 7 → Pred = △

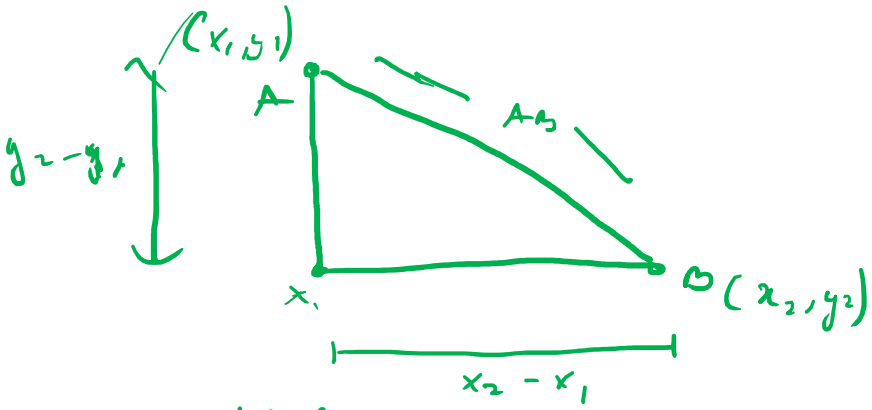
→ Distance =)

① Euclidean distance –

$$H^2 = P^2 + B^2$$

$$AB^2 = (y_2 - y_1)^2 + (x_2 - x_1)^2$$

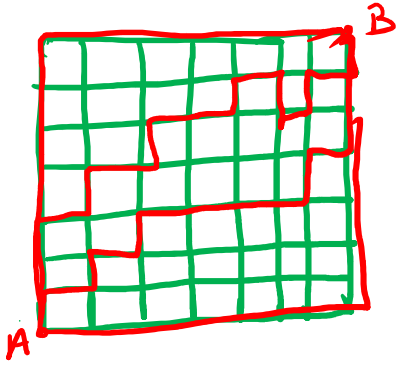
$$AB = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \rightarrow 2\text{-D}$$



– (3-D) $x, y, z, w, q \ldots\ldots F_{100}$

$$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2 + (w_2 - w_1)^2 + (q_2 - q_1)^2 + \ldots + (F_{100_2} - F_{100_1})^2}$$

② Manhattan Distance / city block dist.



$$\Sigma |y_2 - y_1| + |x_2 - x_1| + \ldots\ldots + |F_{100_2} - F_{100_1}|$$

→ How to choose the best value of K → based on performance

→ > 1000 Rows → Thumb Rule → $K = 1, \sqrt{N}$ (lower, N/2)

→ ≤ 1000 Rows → " " $K = 1, N/2$

→ K should not be small

2 classes ⇒ K = odd | K = Even

# of classes even — odd

1 − 100

99

98

1 − $\sqrt{2}$ →

$K = \sqrt{2}$: 99%

$K = \sqrt{1}$ — 94%

$25 - \sqrt{0}$

10%

100

10

K = 2