K NN classification
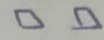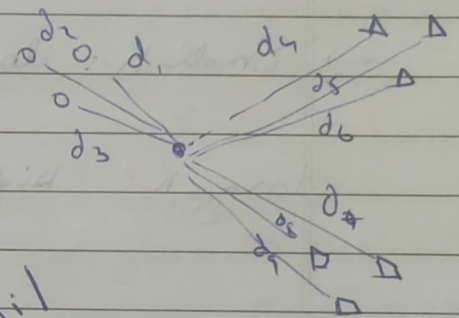
Step 1 Initialization
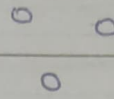
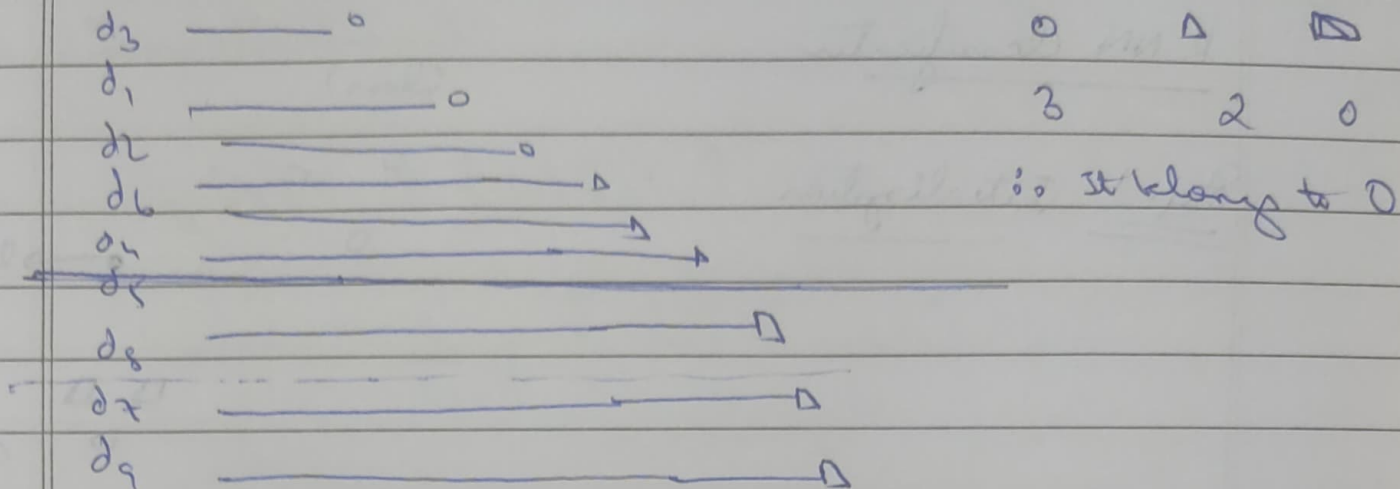**Step 2** Calc distance

Euclidean :

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

Manhattan: $d(x,y) = \sum_{i=1}^{n} |x_i - y_i|$

Calculate distance from given pt to each sample

Step 3 Short distance

Sort the nearest neighbours of the given pt by distance in ↑ order

d3 ——— ○
d1 ——— ○
d2 ——— ○
d6 ——————— △
d4 ——————— △
d5 ——————— △
d8 ——————————— ◻
d7 ——————————— ◻
d9 ——————————— ◻

Step 4 Let's keep k = 5

d_3 ———————o

d_1 ————————o

d_2 ——————————o

d_6 ———————————△

d_4 —————————————▲

d_5 ———————————————

d_8 ————————————□

d_7 ——————————————△

d_9 ———————————————△

	O	△	▱
	3	2	0

∴ It belongs to O

→ Small K : low bias , high variance , overfitting.

large K : High bias , low variance , underfitting

→ K - generally odd no. to minimize ties