Summary of Previous Lecture: (1-NN → Voronoi Diagram)

↳ Non-parametric Learning ⇒ K-Nearest Neighbour

$$\hat{f}(q) = \underset{v \in V}{\operatorname{argmax}} \sum_{i=1}^{K} w_i \, \delta(v, f(x_i))$$

where, V = Set of classes

f = unknown labelling function

$w_i = \frac{1}{d(q,x_i)^2}$

$$\delta(a,b) = \begin{cases} 1, & \text{if } a = b \\ 0, & \text{otherwise} \end{cases}$$

($w_i \Rightarrow$ distance-weighted K-NN)

Distances: — Euclidean $= \sqrt{\sum_{i=1}^{m} (x_i - y_i)^2} \leftarrow D(x,y)$

↳ Normalize

↳ Kernel function (Perzen Window)

Manhattan $= \sum_{i=1}^{m} |x_i - y_i| \leftarrow D(x,y)$

---- etc.

Value of K $= \sqrt{N}$ ($N = \#TE$) or $K = \frac{N}{10}$.

→ Issues: — Curse of Dimensionality
— Computational Complexity

TODAY'S LECTURE

N → TE
D → Attribute

K-NN Algo: (for any point q)

1. ↳ Compute distance of q to all other points
2. ↳ Find top k points
3. ↳ Check for the majority and label q.

▶ **Complexity:**

1. $O(ND)$
2. $O(KN)$
3. $O(K)$

$\Rightarrow O(ND + KN)$

$O(K) + O[(N-K)\log K] = O(N \log K)$

(Optimizations over KNN)

→ K ← $K = \sqrt{N}$, $N/10$ (exp.)

→ N → How to reduce effective TE

→ D → How to represent succinctly

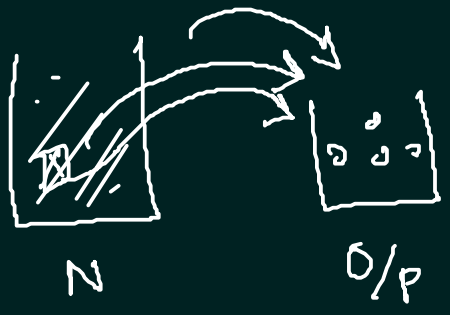
▷ How can we select Subset of N points such that the Decision Boundary is unaltered?

def1 → Decision Boundary Consistent

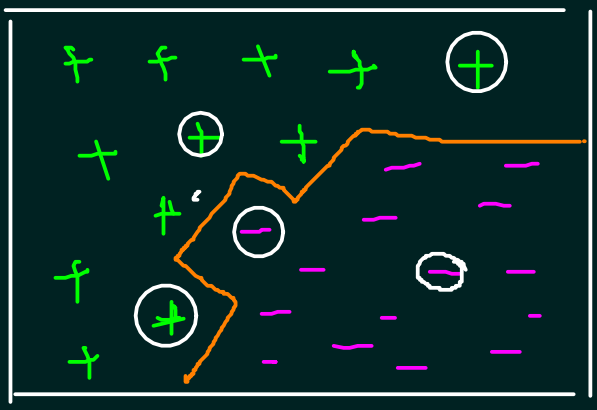
def2 → Minimum Consistent Set (condensing)

?

▷ Condensing:



$({}^{N}C_{2} * N)$

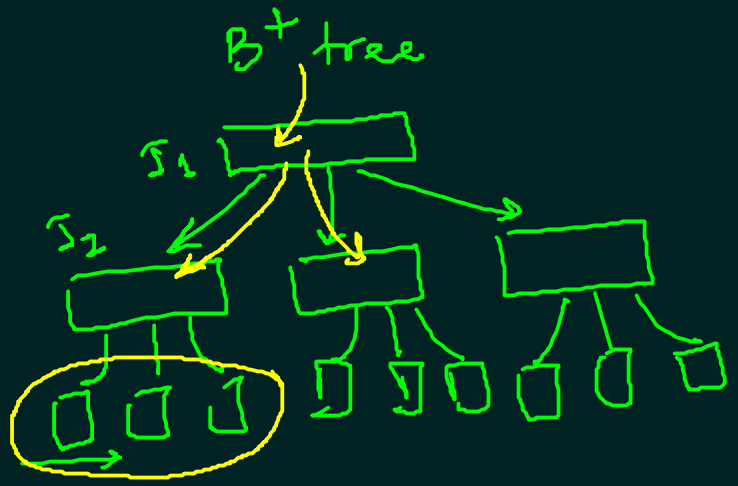


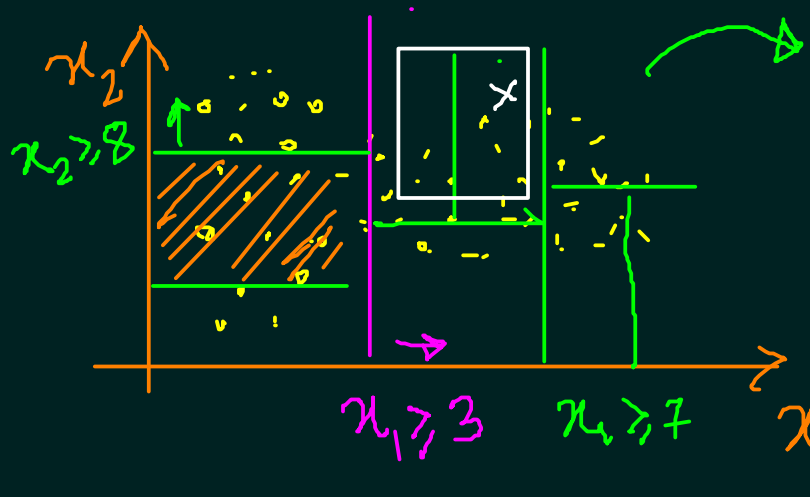
→ Algo. Complexity: $O(N^3)$ ← (worst)

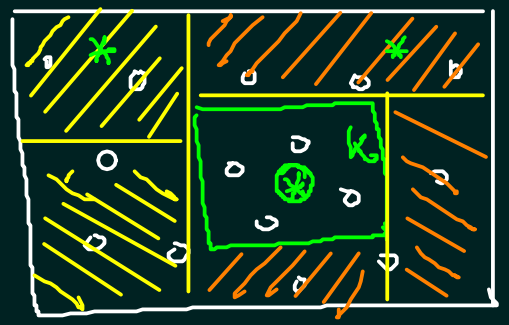
→ choosing a misclassified point

→ Computing the boundary (after that misclassified point put into final set)

▷ How to represent Dimensionality of attributes?



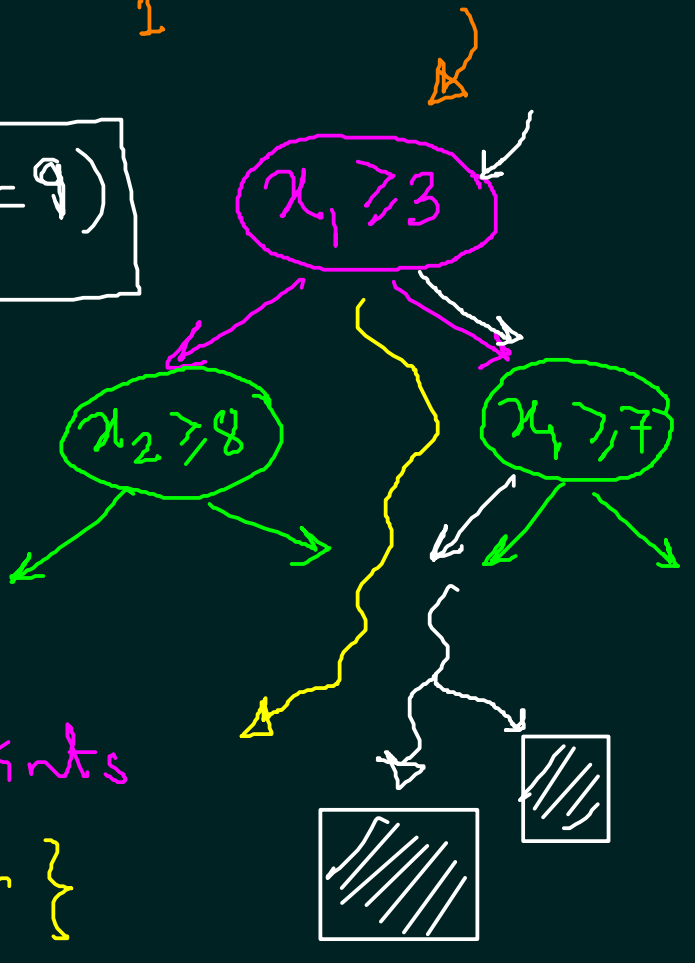




K-d Trees

$q = (x_1 = 5, x_2 = 9)$

Apply



▷ K-d Trees → efficient in searching for new points

→ Reduced N

Reduced dimensionality → Index term }

Application:

- Medical Domain Diagnostics } Instances to look for (TE) (Instance-based Learning)
- Judgement / Law Verdict } Cases prior to this (TE) (Case-based Learning)
- Feature of k-NN: lazily find out nearest neighbours
  - Lazy Learning (greedy classification)