KNN:-

Finding the k nearest neighbors of q.

Distance Metric:-

To determine closeness, k-NN uses a distance metric.

1) Euclidean Distance:-

$$d(x_i, q) = \sqrt{\sum_{j=1}^{d} (x_{ij} - q_j)^2}$$

2) Manhattan Distance:-

$$d(x_i, q) = \sum_{j=1}^{d} |x_{ij} - q_j|$$

3) Minkowski Distance:-

$$d(x_i, q) = \left(\sum_{j=1}^{d} |x_{ij} - q_i|^p \right)^{1/p}$$

→ for $p = 2$, it becomes Euclidean distance

→ for $p = 1$, it becomes Manhattan distance.

Classification with K-NN:

1) Finding Neighbors:-

→ Compute the distance $d(x_i, q)$ for all data points in the dataset.

2) Majority voting:-

$\hat{y}$ = mode $\{y_{i_1}, y_{i_2}, \ldots, y_{i_k}\}$.

Regression with k-NN:-

1) finding neighbors

2) Prediction,

$$\hat{y} = \frac{1}{k} \sum_{i \in \text{neighbors}} y_i$$

k-NN Algorithm:-

Prediction phase:-

1) Compute distances.

2) Sort distances & select k closest-

neighbors.

3) Aggregate the neighbors' labels. or values.

Choosing k:-

1) Small k

2) Large k

→ Common choice → $k = \sqrt{n}$