

# INTERVIEW QUESTIONS

## KNN:

### 1. What are the key hyperparameters in KNN?

The key hyperparameters in K-Nearest Neighbors (KNN) include:

- **K (Number of Neighbors):** This is the most important hyperparameter. It determines how many neighbors are considered when making a prediction. A small value of K can be noisy and sensitive to outliers, while a large value can smooth out the prediction but may ignore local patterns.
- **Distance Metric:** This defines how the distance between data points is calculated. Common choices include Euclidean distance, Manhattan distance, and Minkowski distance.
- **Weights:** This parameter determines whether all neighbors contribute equally to the prediction or if closer neighbors have more influence. Options include uniform weights (all neighbors are weighted equally) and distance-based weights (closer neighbors have greater influence).
- **Algorithm:** This specifies the method used to compute the nearest neighbors, such as 'auto', 'ball\_tree', 'kd\_tree', or 'brute'.
- **Leaf Size:** This is relevant when using tree-based algorithms for speeding up the search for neighbors. It defines the number of points in a leaf node and can impact performance.

### 2. What distance metrics can be used in KNN?

Several distance metrics can be used in KNN, including:

- **Euclidean Distance:** The most commonly used metric, it calculates the straight-line distance between two points in Euclidean space.
- **Manhattan Distance:** Also known as L1 distance or taxicab distance, it measures the absolute differences along each dimension.
- **Minkowski Distance:** A generalization of both Euclidean and Manhattan distance, defined using a parameter  $p$ . For  $p=1$ , it's Manhattan distance; for  $p=2$ , it's Euclidean distance.
- **Hamming Distance:** Useful for categorical data, it measures the proportion of differing attributes between two binary vectors.
- **Cosine Similarity:** Although not a distance metric per se, it measures the cosine of the angle between two vectors and is often used for text data.
- **Chebyshev Distance:** Also known as maximum metric, it measures the greatest of the differences along any coordinate dimension.

