1. Can we use KNN for Regression?

Yes, KNN can be used for regression algorithm as well.

K-Nearest Neighbours (KNN) is a non-parametric machine learning algorithm that can be used for both classification and regression tasks. In the context of regression, KNN is often referred to as "K-Nearest Neighbours Regression" or "KNN Regression." It's a simple and intuitive algorithm that makes predictions by finding the K nearest data points to a given input and averaging their target values (for numerical regression) or selecting the majority class (for classification).

Here's an overview of how KNN Regression works:

Data Collection: You start with a dataset that includes both input features and target values. In regression tasks, the target values are continuous and represent the output you want to predict.

Choosing the Number of Neighbours (K): You need to choose the number of nearest neighbours, K, that will be used to make predictions. This is a hyperparameter that you can tune based on the characteristics of your data. A small K (e.g., 1 or 3) may lead to noisy predictions, while a large K may lead to overly smoothed predictions.

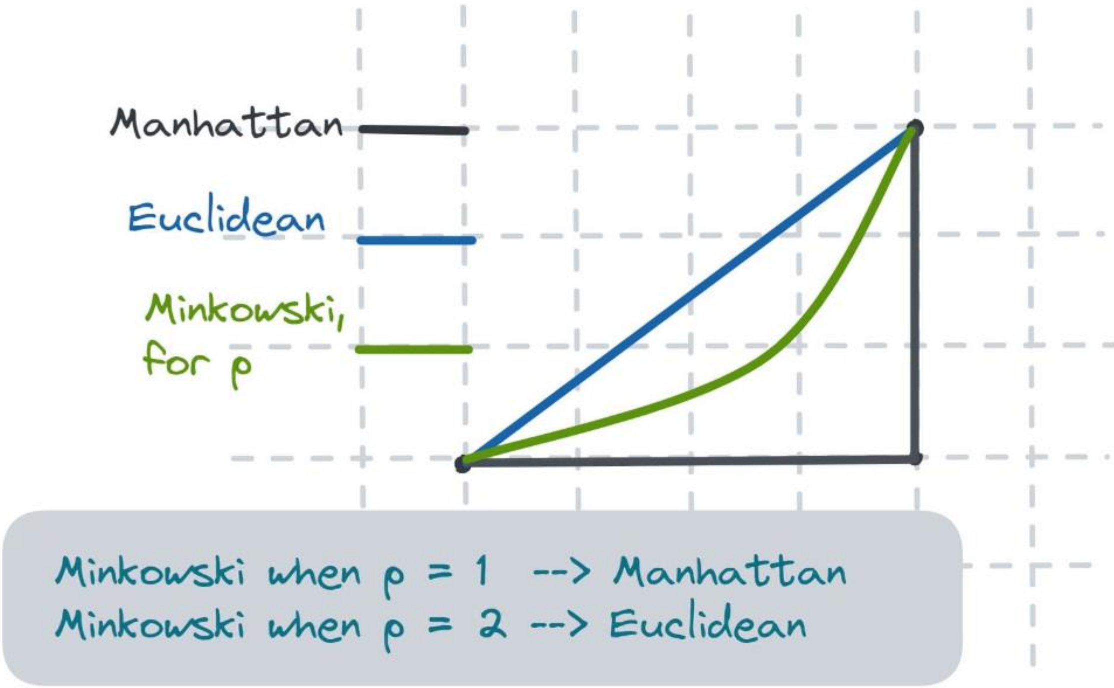
Distance Metric: KNN relies on a distance metric (e.g., Euclidean distance) to measure the similarity between data points. Different distance metrics can be used depending on the nature of your data.

Prediction: When you want to make a prediction for a new input data point, KNN calculates the distance between this point and all other data points in the dataset. It then selects the K data points with the smallest distances.

Regression Prediction: For regression, the predicted value for the new data point is the average of the target values of the K nearest neighbours. This could be a simple arithmetic mean.

In other words, the KNN algorithm can be applied when the dependent variable is continuous. For regression problem statements, the predicted value is given by the average of the values of its k nearest neighbours.

2. Alternative for Euclidean distance:



Euclidean - Euclidean distance is the shortest distance between any two points in a metric space.

$$d(\mathbf{x},\mathbf{y}) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$$

Manhattan - Manhattan Distance is the sum of absolute differences between points across all the dimensions.

$$d(\mathbf{x},\mathbf{y}) = |x_1 - y_1| + |x_2 - y_2|$$

Minkowski - Generalization of Euclidean and Manhattan distance.

$$d(\mathbf{x},\mathbf{y}) = \left(\sum_{i=1}^{n} |x_i - y_i|^p\right)^{1/p} \quad for\ p \geq 1$$