What is K-NN Regression?

K-Nearest Neighbors (KNN) Regression is a **supervised machine learning algorithm** used to predict **continuous numeric values**. It works by finding the **K closest data points** (neighbors) to a new input and predicts the output value based on the **average** (or sometimes weighted average) of those neighbors' values.

It’s simple, effective, and requires **no training** — the algorithm makes predictions based on raw data during testing time.

Basic Idea

Imagine you want to guess the price of a house. You look at the prices of the **K most similar houses nearby**. Then, you take the **average** of those prices — that’s your predicted price.

That's how KNN regression works: it trusts nearby data values.

How does K-NN Regression work?

The K-NN Regression working can be explained using this step-by-step guide:

Step 1: Select the number **K** of the neighbors (odd/even).

Step 2: Calculate the **distance** (typically Euclidean) between the new data point and all existing data points in the dataset.

Step 3: Take the **K nearest neighbors** as per the calculated distances. Sort the distances in ascending order.

Step 4: Find the **target values** (outputs) of these K neighbors.

Step 5: Take the **mean** (or weighted average) of these K values.

Step 6: Assign this average as the **predicted value** for the new point.

Quick Example

Let’s say you want to predict the **price of a house** based on area:

Area (sqft)	Price ($k)
1000	200
1100	210
1200	220
1300	230
1400	240

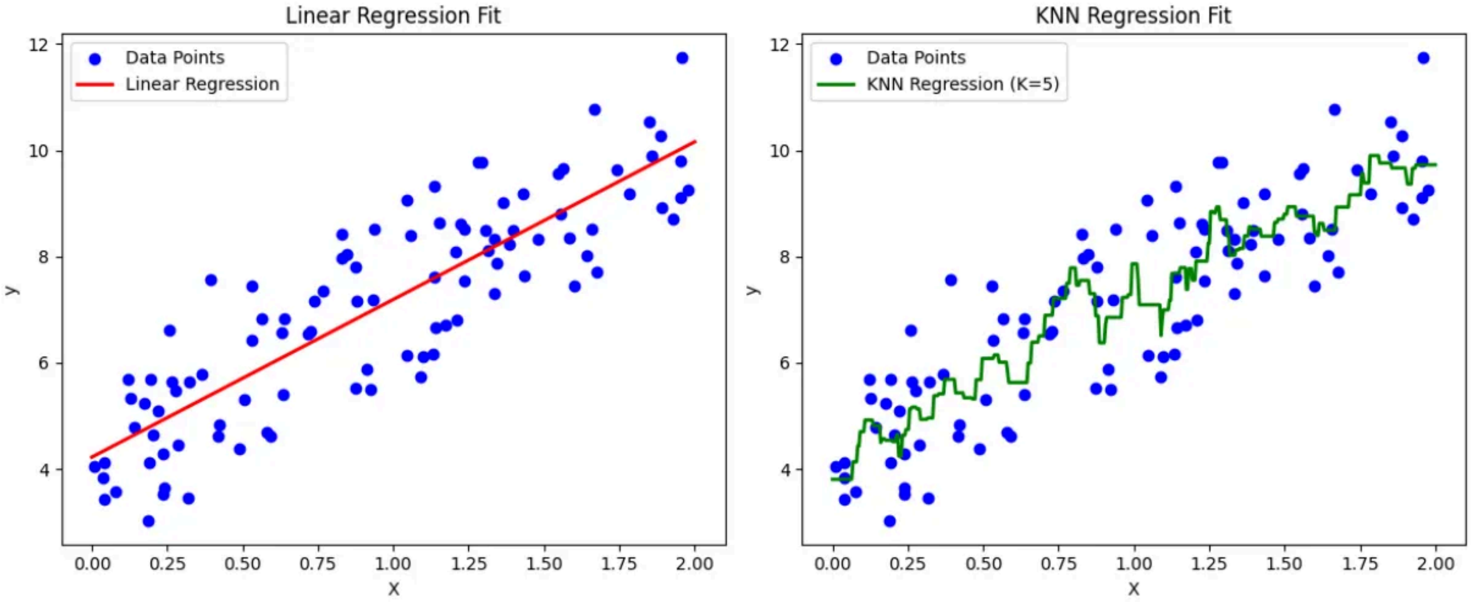
Now we have a new house: **1150 sqft**. Choose K = 3.

- The 3 closest areas: 1100, 1200, 1000
- Corresponding prices: 210, 220, 200
- Average = (210 + 220 + 200) / 3 = **210**

Predicted Price = **$210k**

Things to Keep in Mind

- **K matters**: Too small = noise, too big = over smoothing. A good starting point is 5% of the total dataset.
- **Feature scaling**: Normalize your data, especially when features have different units.
- **Weighted KNN**: Give closer neighbors more influence.
- **Distance metrics**: Try others like Manhattan, Minkowski depending on the data.
- **Curse of Dimensionality**: KNN doesn’t perform well with too many features.


Linear Regression Fit
Data Points
Linear Regression
12
10
8
6
4
y
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
X
KNN Regression Fit
Data Points
KNN Regression (K=5)
12
10
8
6
4
y
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
X