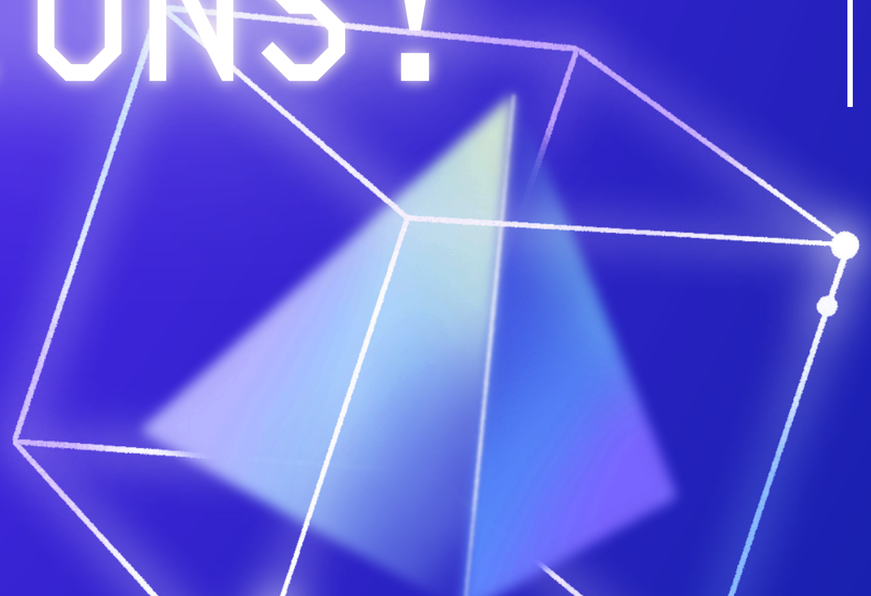
AI/MACHINE LEARNING WORKSHOP

DAY 9: INTRODUCTION TO AI & ML: REGRESSION K-NEAREST NEIGHBOR

Youth Oppurtunties in Tech Inovation

REMINDER PLEASE ASK QUESTIONS!

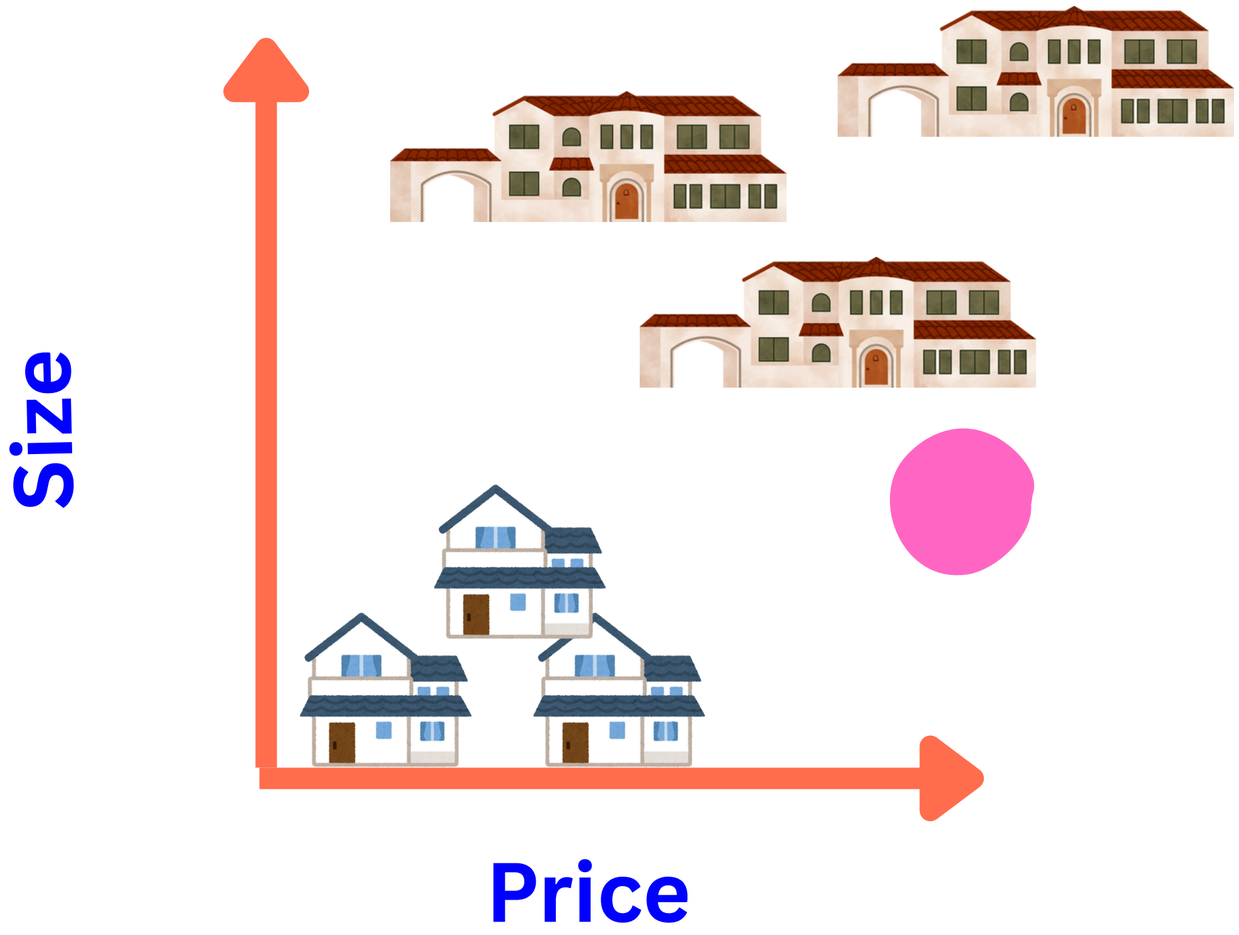


WHAT IS K-NEAREST NEIGHBOR

SUPERVISED MACHINE LEARNING ALGORITHM THAT CLASSIFIES OR PREDICTS THE VALUE OF A NEW DATA POINT BASED ON THE MAJORITY CLASS (OR AVERAGE VALUE) OF ITS "K" NEAREST NEIGHBORS IN THE TRAINING DATA

KNN Model

KNN Model



KNN Model

Similair data points are located near each other

KNN Model

Step-by-Step: How KNN Works

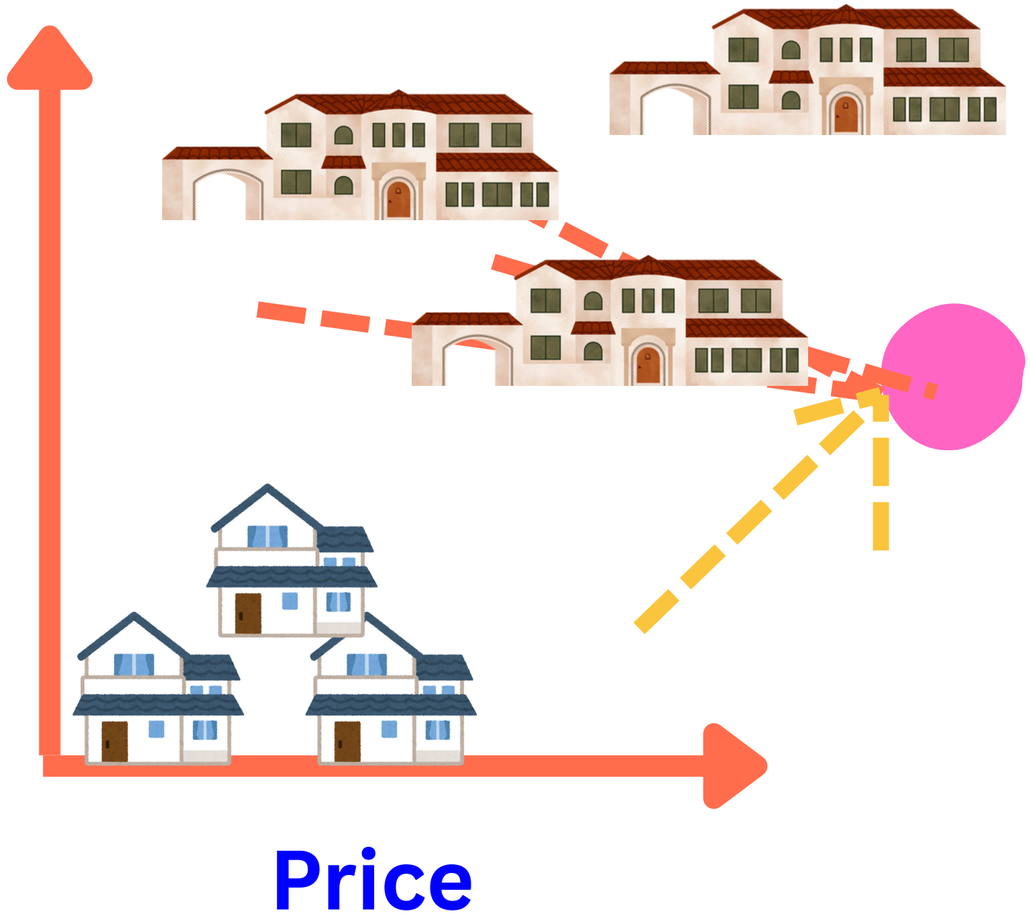
- **Choose a value for k (number of neighbors).**
- **Calculate the distance from the new point to all other points in the dataset.**
- **Sort distances and pick the k nearest points.**
- **average (regression) to make a prediction.**

KNN Model

Calculate the Distance: Euclidean Distance

- New input: $\vec{x} = (x_1, x_2, ..., x_n)$
- Training point: $\vec{x}^{(i)} = (x_1^{(i)}, x_2^{(i)}, ..., x_n^{(i)})$

$$d(\vec{x}, \vec{x}^{(i)}) = \sqrt{\sum_{j=1}^{n} \left(x_j - x_j^{(i)}\right)^2}$$



KNN Model

Sort the Distances

D1

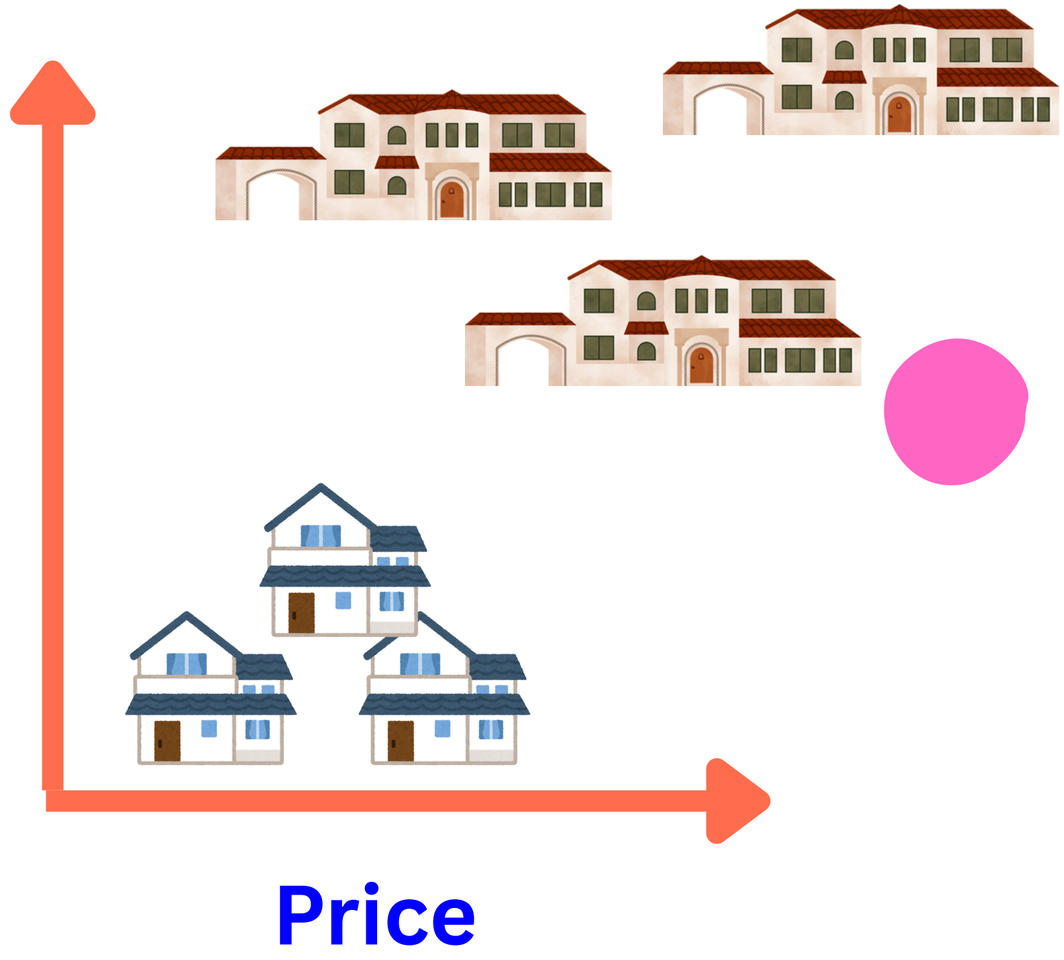
D2

D3

D4

D5

D6



KNN Model

Sort the Distances

D1

D2

D3

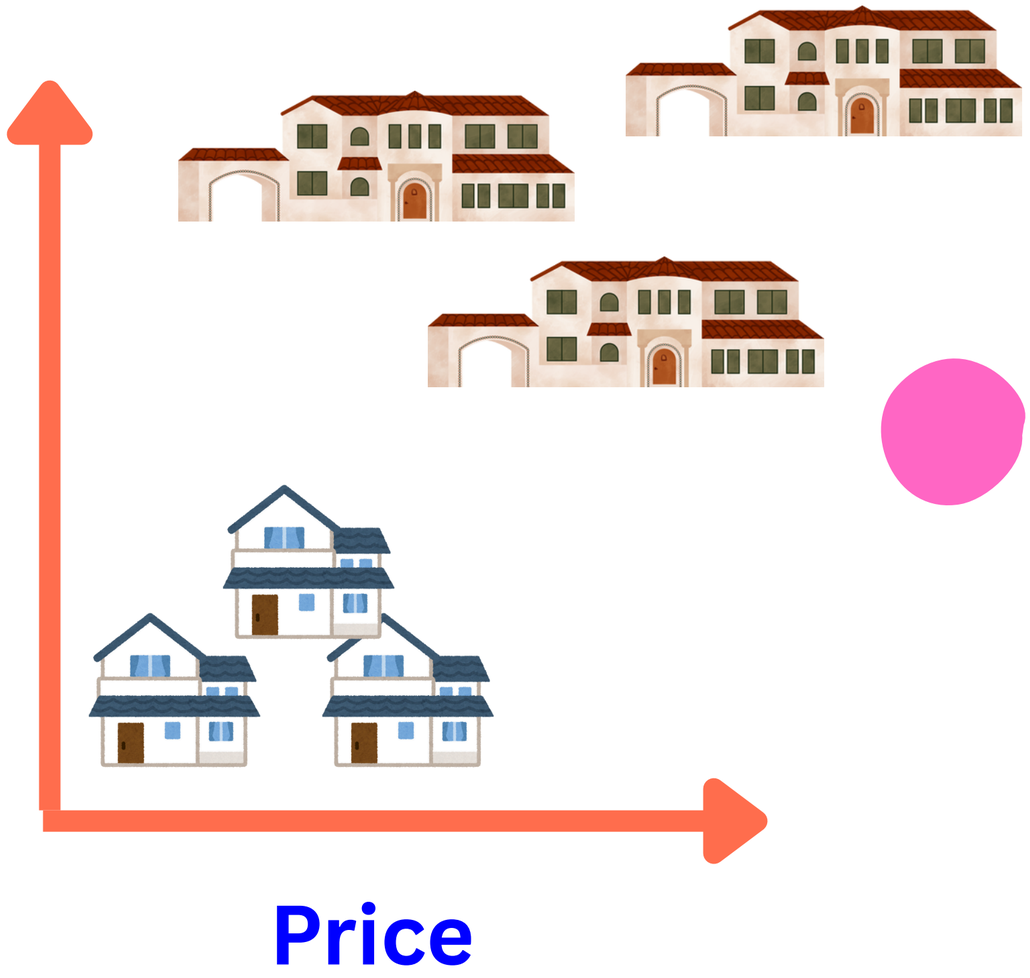
D4

D5

D6

Calculate the Regression

$$\hat{y} = \frac{1}{k} \sum_{j=1}^{k} y_{i_j}$$



KNN Model

Size (sqft)	Price ($1000s)
1000	200
1200	240
1300	260

You want to predict the price for a 1250 sqft house using $k = 2$.

Step 1: Compute distances:

- $d(1250, 1000) = 250$
- $d(1250, 1200) = 50$
- $d(1250, 1300) = 50$
- $d(1250, 1500) = 250$

Step 2: Nearest 2 neighbors: 1200 (240) and 1300 (260)

Step 3: Average:

$$\hat{y} = \frac{240 + 260}{2} = 250$$

QUESTIONS AND FINAL THOUGHTS!

THANK YOU!