Foundations Of Neural Networks and Deep Learning

Day-3

SME: Gagan P

Contact: gaganp000999@gmail.com

recap:

1. In the complete machine learning pipeline, what is the first step after data collection?

A. Splitting the data
B. Preprocessing and cleaning the data
C. Feature engineering
D. Model training and evaluation

recap:

1. In the complete machine learning pipeline, what is the first step after data collection?

A. Splitting the data
B. Preprocessing and cleaning the data
C. Feature engineering
D. Model training and evaluation

Which of the following is an example of an outlier in a dataset?

A. A house price of $750,000 in a neighborhood where prices range from $700,000 to $800,000.

B. A house price of $20 million in a neighborhood where prices are typically between $500,000 and $1 million.

C. A house with three rooms and an area of 1,200 sq. ft.

D. A duplicated entry in a dataset.

Which of the following is an example of an outlier in a dataset?

A. A house price of $750,000 in a neighborhood where prices range from $700,000 to $800,000.

B. A house price of $20 million in a neighborhood where prices are typically between $500,000 and $1 million.

C. A house with three rooms and an area of 1,200 sq. ft.

D. A duplicated entry in a dataset.

What is a "feature" in a machine learning dataset?

A. The predicted output or label.
B. The measurable property used as an input to the model.
C. A type of machine learning algorithm.
D. The final trained model.

What is a "feature" in a machine learning dataset?

A. The predicted output or label.

B. The measurable property used as an input to the model.

C. A type of machine learning algorithm.

D. The final trained model.

What does the notation $x^{(i)}$ represent?

A. The value of the i-th feature.

B. The value of the i-th label.

C. The feature vector of the i-th sample.

D. The i-th dimension of the dataset.

What does the notation $x^{(i)}$ represent?

A. The value of the i-th feature.

B. The value of the i-th label.

C. The feature vector of the i-th sample.

D. The i-th dimension of the dataset.

For a house price prediction model, which of the following is considered the target or label, represented by $y^{(i)}$?

A. Number of bedrooms
B. House age
C. Hot water availability
D. The house price

For a house price prediction model, which of the following is considered the target or label, represented by $y^{(i)}$?

A. Number of bedrooms
B. House age
C. Hot water availability
D. The house price

Which of the following is a reason for performing feature engineering?

A. To increase the number of missing values.

B. To ensure all features have different scales.

C. To prevent larger-scale features from dominating the learning process.

D. To remove all features from the dataset.

Which of the following is a reason for performing feature engineering?

A. To increase the number of missing values.

B. To ensure all features have different scales.

C. To prevent larger-scale features from dominating the learning process.

D. To remove all features from the dataset.

What is the formula for Normalization (Min-Max Scaling)

A. $x_{std} = (x - \mu)/\sigma$

B. $x_{norm} = (x - x_{min})/(x_{max} - x_{min})$

C. $x_{norm} = (x - x_{max})/(x_{min} - x_{max})$

D. $x_{std} = (x - \sigma)/\mu$

What is the formula for Normalization (Min-Max Scaling)

A. $x_{std} = (x - \mu)/\sigma$

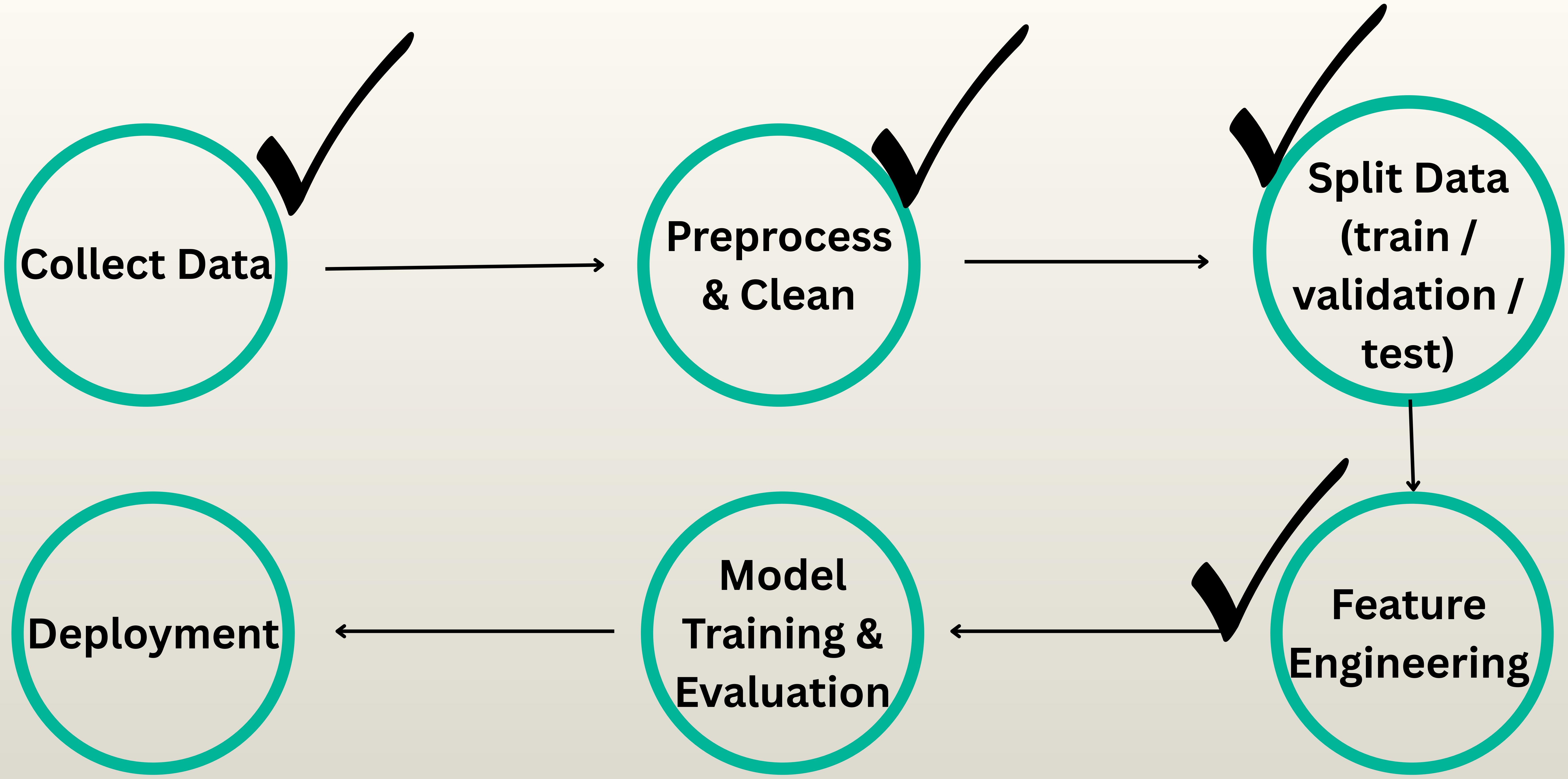
B. $x_{norm} = (x - x_{min})/(x_{max} - x_{min})$

C. $x_{norm} = (x - x_{max})/(x_{min} - x_{max})$

D. $x_{std} = (x - \sigma)/\mu$

day 3 - introduction to linear regression

in the last class we learnt about the first 3 steps of machine learning pipeline


Collect Data
Preprocess & Clean
Split Data (train / validation / test)
Feature Engineering
Model Training & Evaluation
Deployment

this class we will learn about the main step!!

Training The Model

what exactly does training a model mean??

- **observing the data**
- **finding the best ML algorithm which can represent the output in terms of features**
- **finding the parameters that will give the most accurate predictions**

this class we will look at the algorithm:

Linear Regression

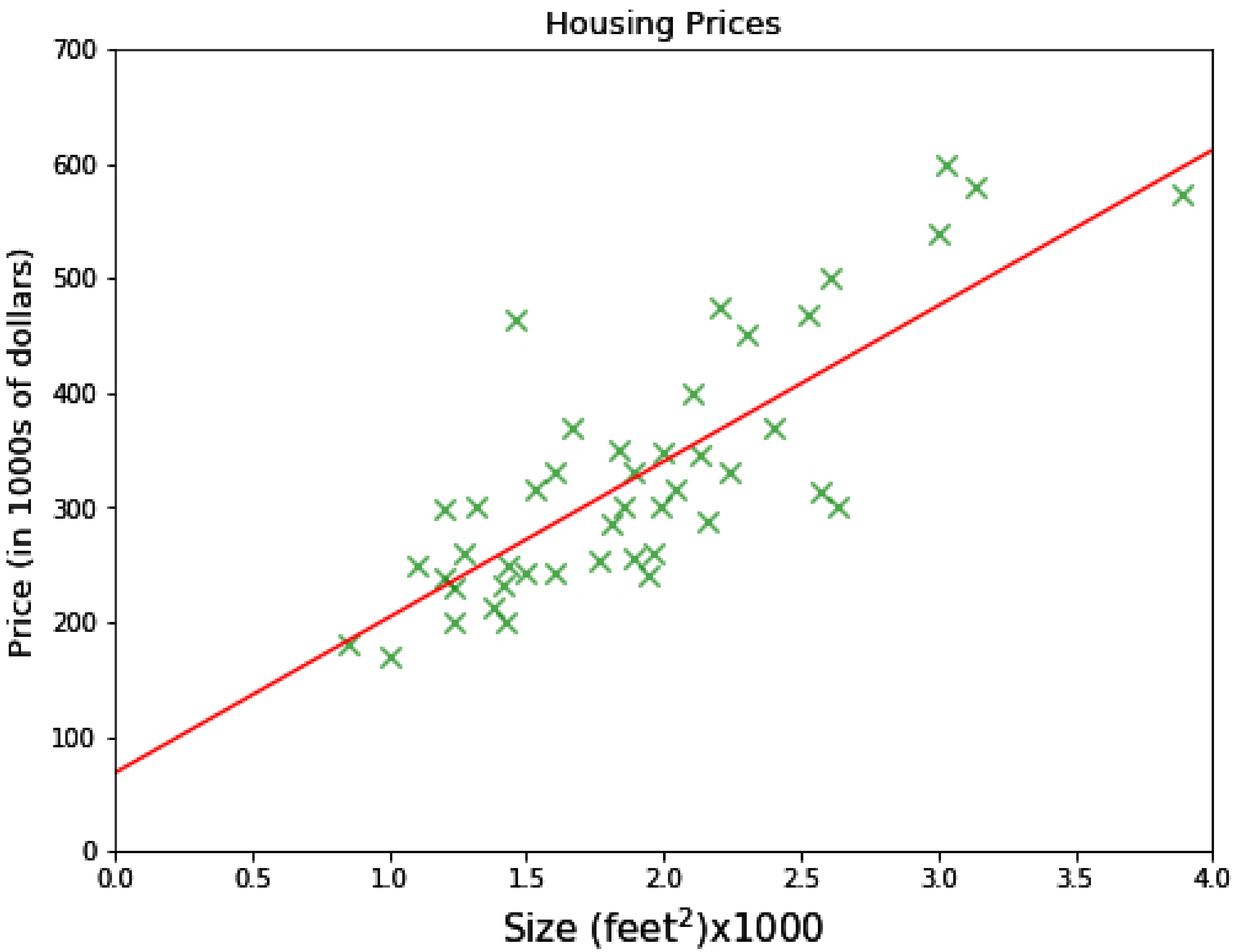
What is Linear Regression?

- A **supervised machine learning algorithm** used for predicting a continuous output (e.g., price, temperature) based on input features.

- The goal is to model the relationship between a dependent variable and one or more independent variables by fitting a **linear equation** to the observed data.

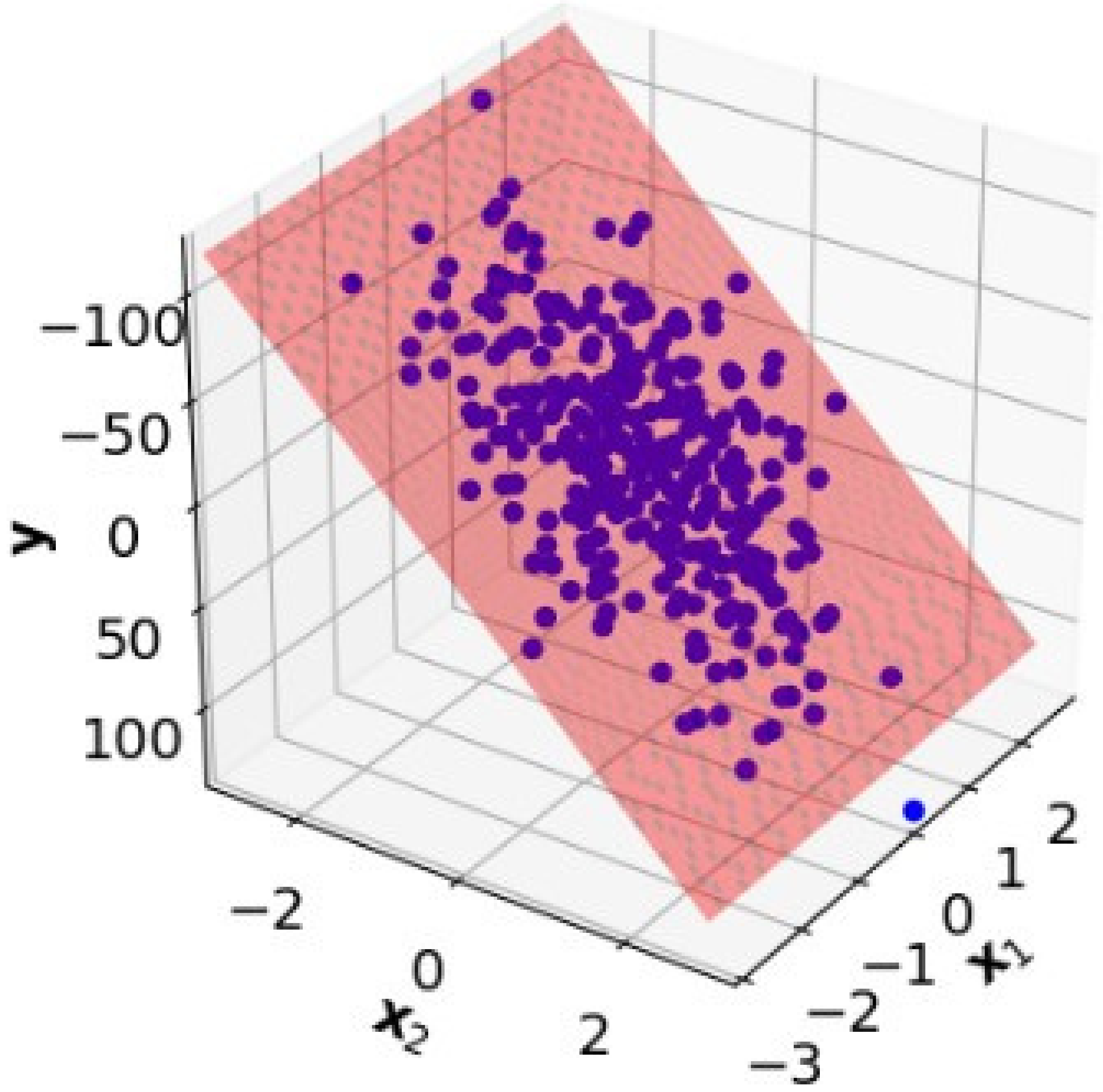
- Analogy: Finding the **best straight line/hyperplane** that passes through a set of data points to represent their relationship.

price in terms of area

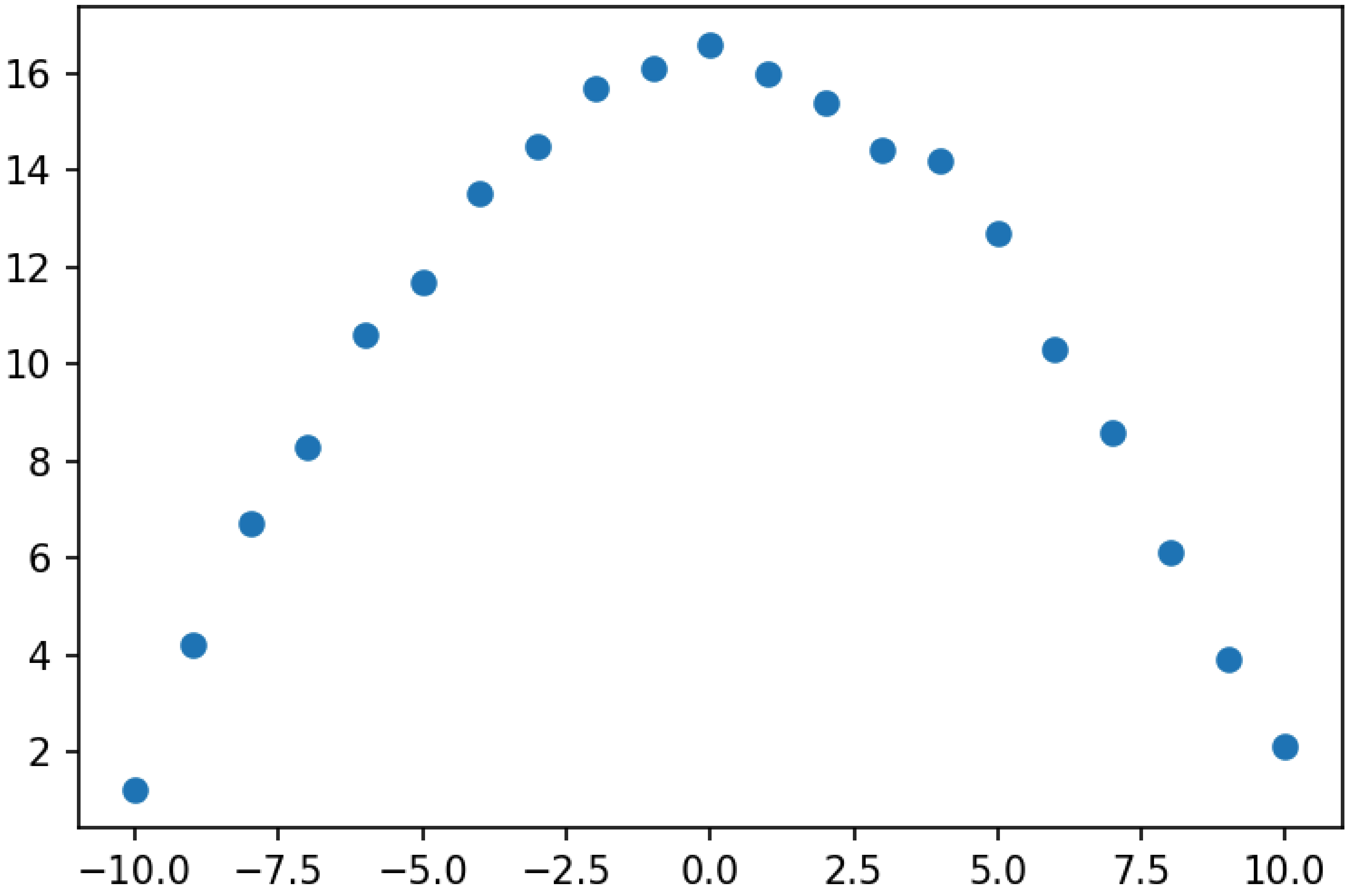


what if we used more than 1 feature to represent the output?

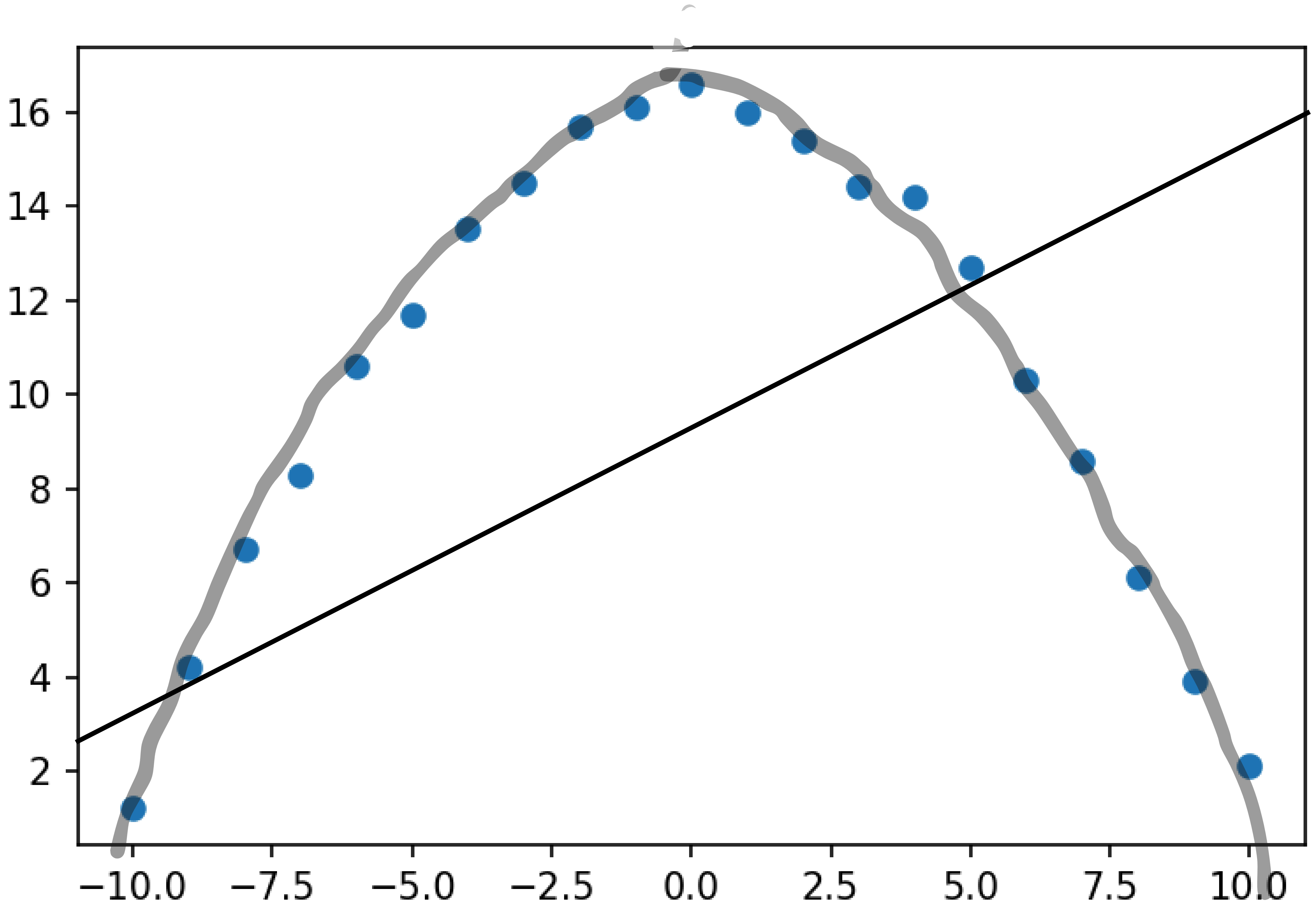
a hyperplane!



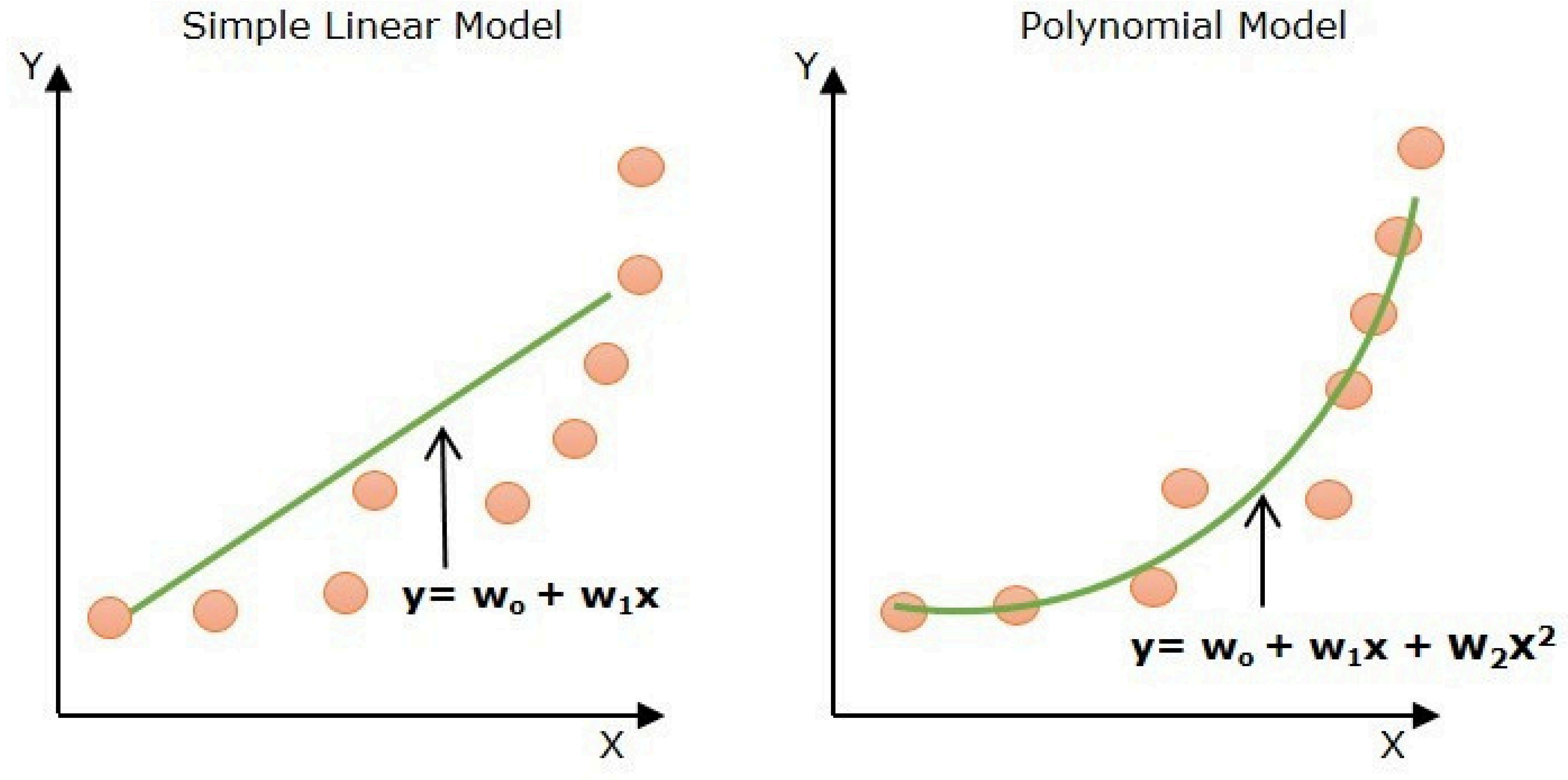
can we apply linear regression here?



can we apply linear regression here?



$$y = -ax^2 + bx + c$$



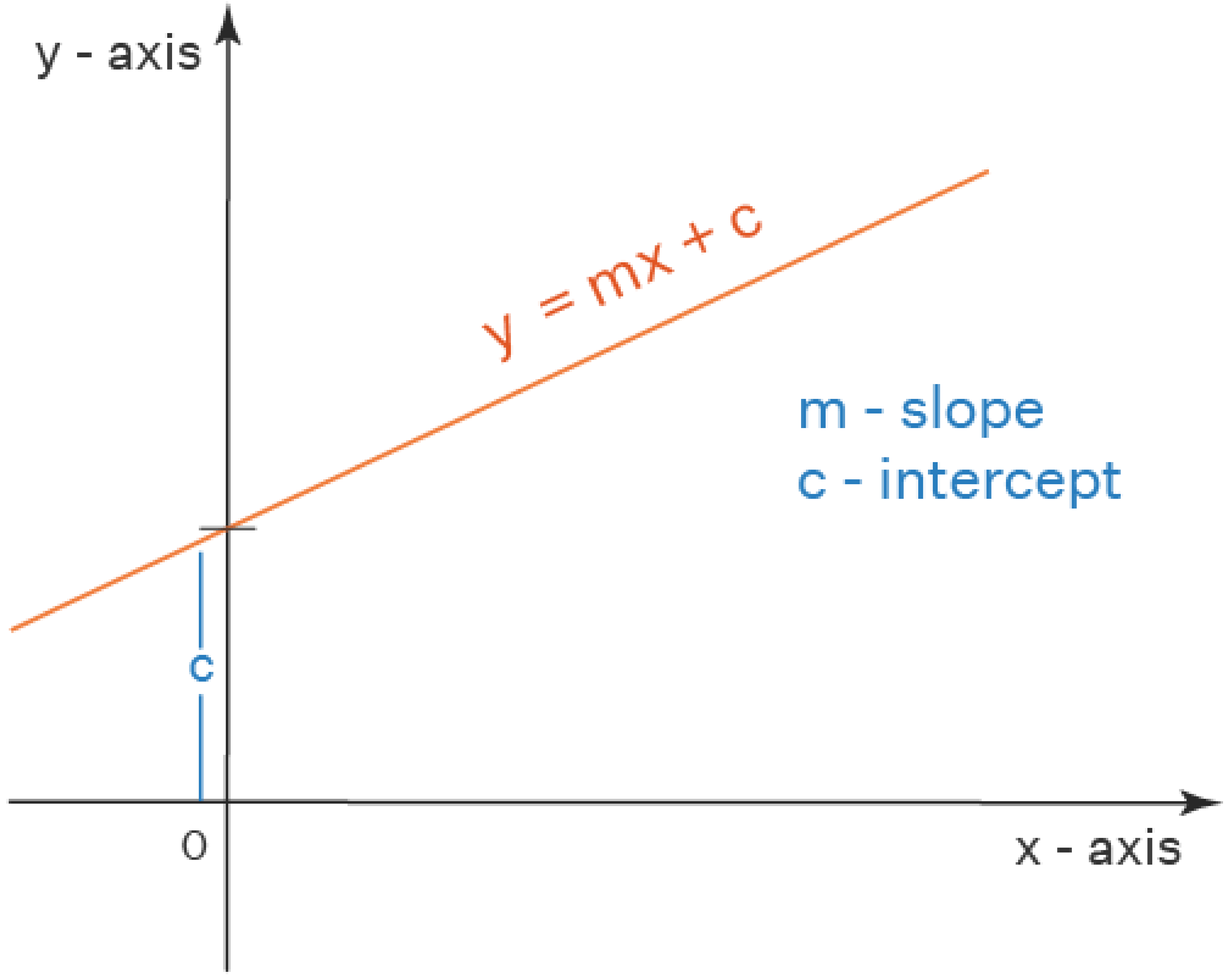
Simple Linear Model
Y
X
y= w₀ + w₁x
Polynomial Model
Y
X
y= w₀ + w₁x + W₂x²

whats the general equation for a straight line?

$$y = mx + c$$

slope

intercept


y - axis
y = mx + c
m - slope
c - intercept
c
O
x - axis

similarly in linear regression we represent the output in terms of input features

$$h(x) = w \cdot x + b$$ → **hypothesis function**

(or)

$$y = \underbrace{w}_{\text{weights}} \cdot \underbrace{x}_{\text{feature}} + b \rightarrow \text{bias}$$

for example x is the area of the house :

$$y(\text{price}) = 2000x + 5000$$

if $x=5000$

$y \sim 1$ cr

linear regression

single

$$y = w \cdot x + b$$

output dependent on single feature

multi

$$y = w_1 x_1 + w_2 x_2 + \cdots + w_n x_n + b$$

output dependent on multiple features

general form:

$$y = \mathbf{w}^T \mathbf{x} + b$$

where:

$\mathbf{w} = [w_1, w_2, \ldots, w_n]$ ⟶ **weight vector**

$\mathbf{x} = [x_1, x_2, \ldots, x_n]$ ⟶ **feature vector**

how do we measure error?

cost/loss function

- A cost function measures the **average error** of a model across all training examples. It tells us how far predictions are from actual values.
- It gives a single number that quantifies how good/bad the model is, so optimization algorithms (like gradient descent) know what to minimize
- there are different types of cost functions
- one of them is **Mean Squared Error or MSE**

Mean Squared Error(MSE)

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - (w^T x_i + b))^2$$

n = number of data points

y_i = true value (actual output)

$\hat{y}_i$ = predicted value from your hypothesis

$y_i - \hat{y}_i$ = error (residual)

a model is better if it has??

- more loss
- ✓ less loss

if the loss is too much

the model should?

- adjust its weights and bias, to decrease the loss
- keep the weights and bias same

our goal: to get the minimum possible loss

↓

to get minimum loss we need to update weights and bias until loss reaches global minima

↓

how does the model know how to update weights?

↓

an algorithm called Gradient Descent(in next class)

hands on session