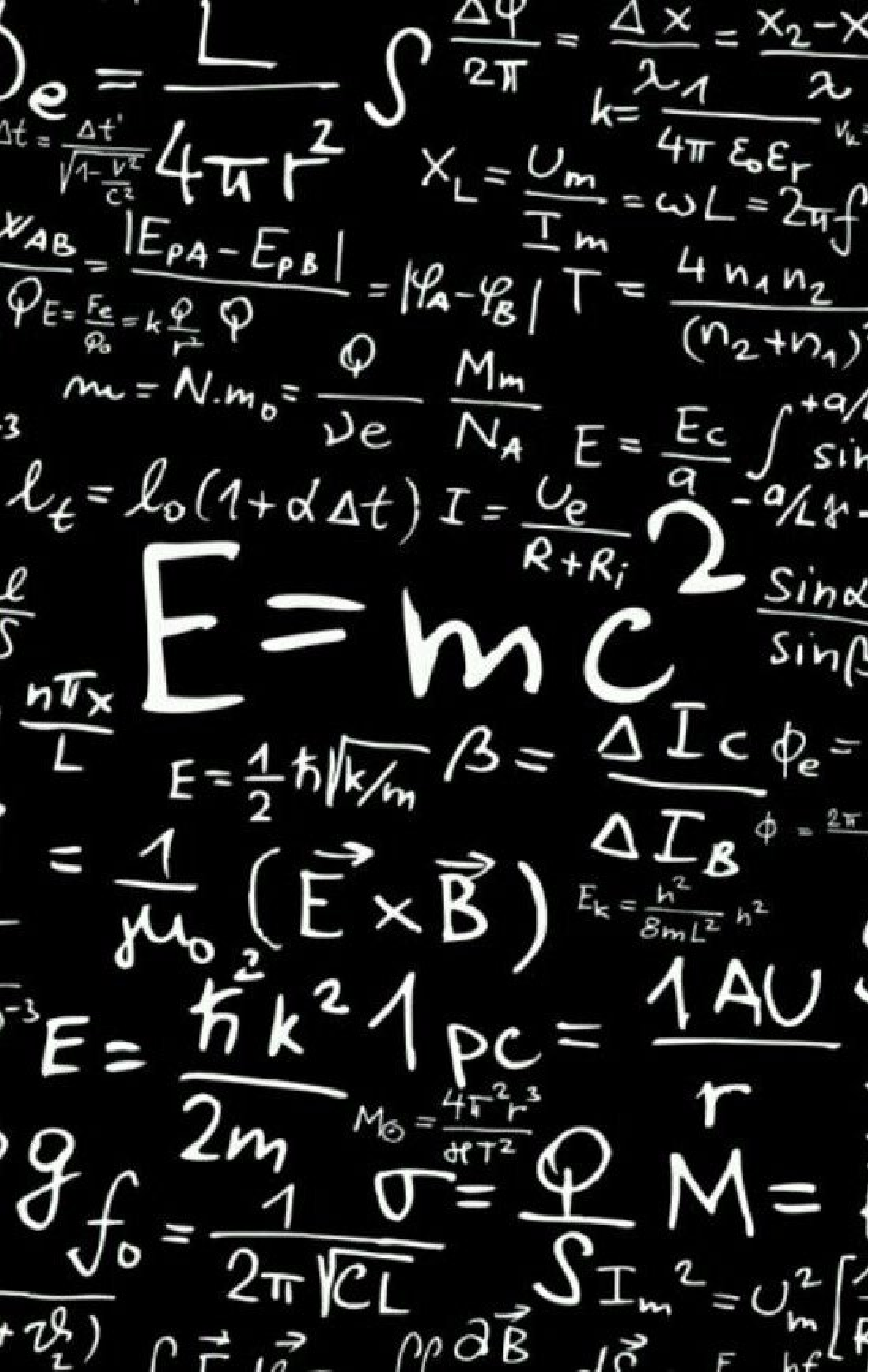


Artificial Intellegence and Machine Learning

Artificial Intelligence

The science of making a Computer Agent that Acts Rationally

Acting Rationally

- Acting Implicitly based on Logic or Reason.

Explicit (adj.)

Tell directly, clearly, and in detail

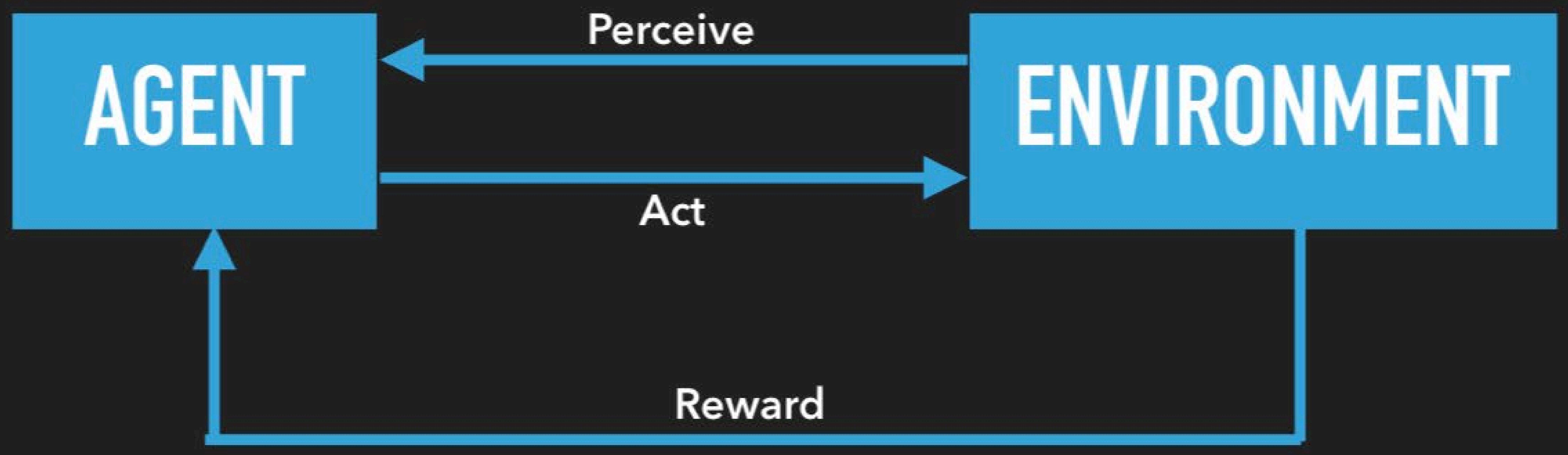
- Ex. If you see a red sign, turn left
- Most of your programming experience are in this space.

Implicit (adj.)

Imply though not plainly expressed

- Not telling the condition directly
- State the conditions as a logic
- Ex. Best route is the shortest route
- In AI/ML, we will play in this space

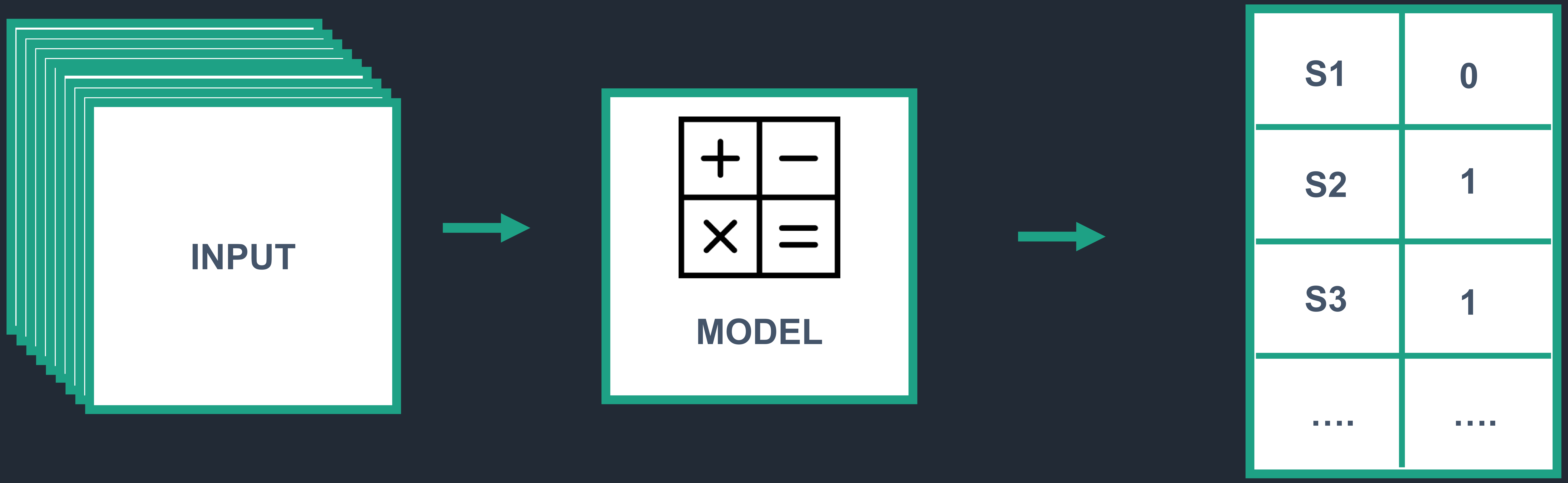
AN INTELLIGENT AGENT



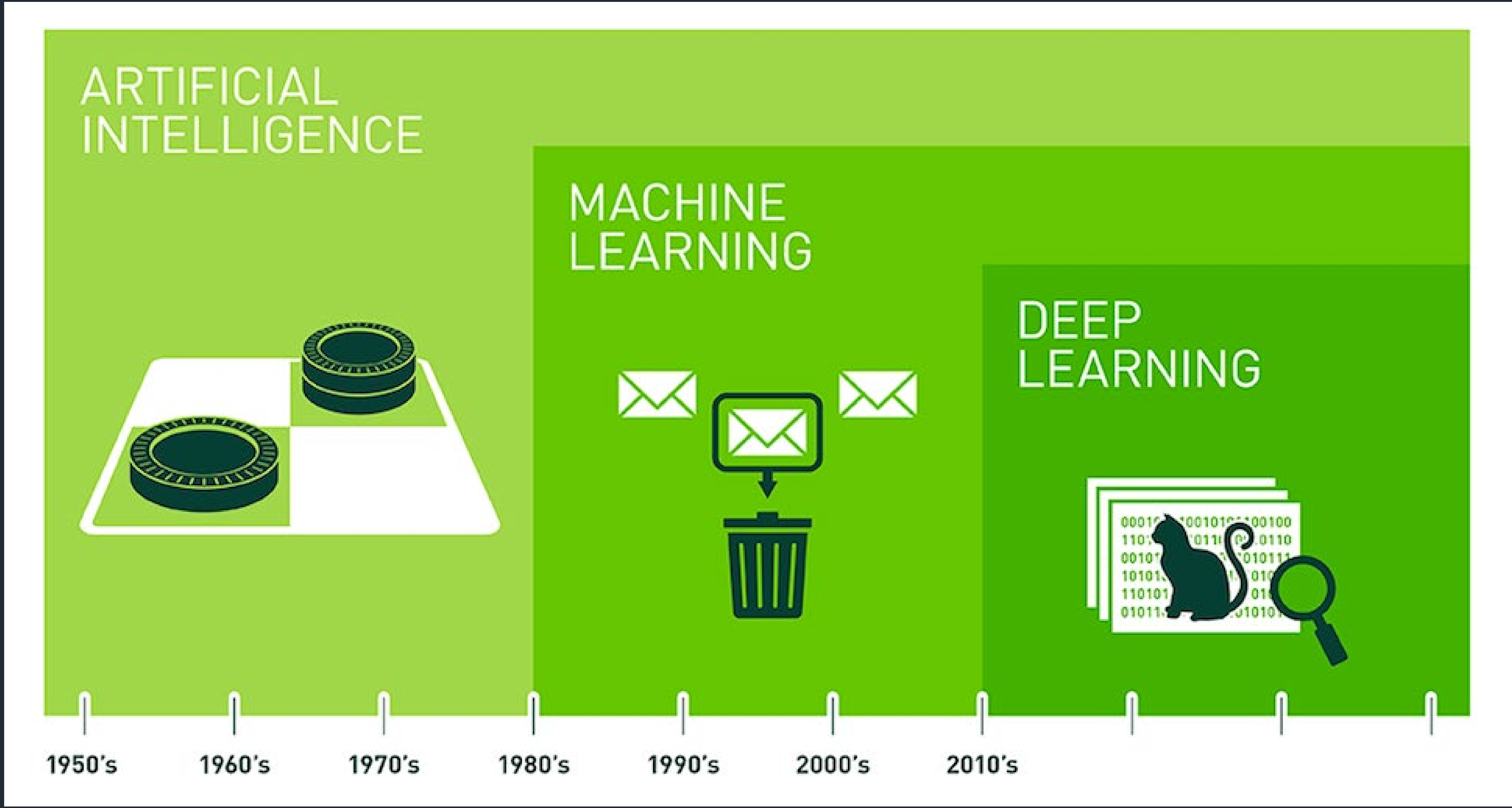
machine learning

The science of getting computers to learn from data without having to be explicitly programmed by humans.

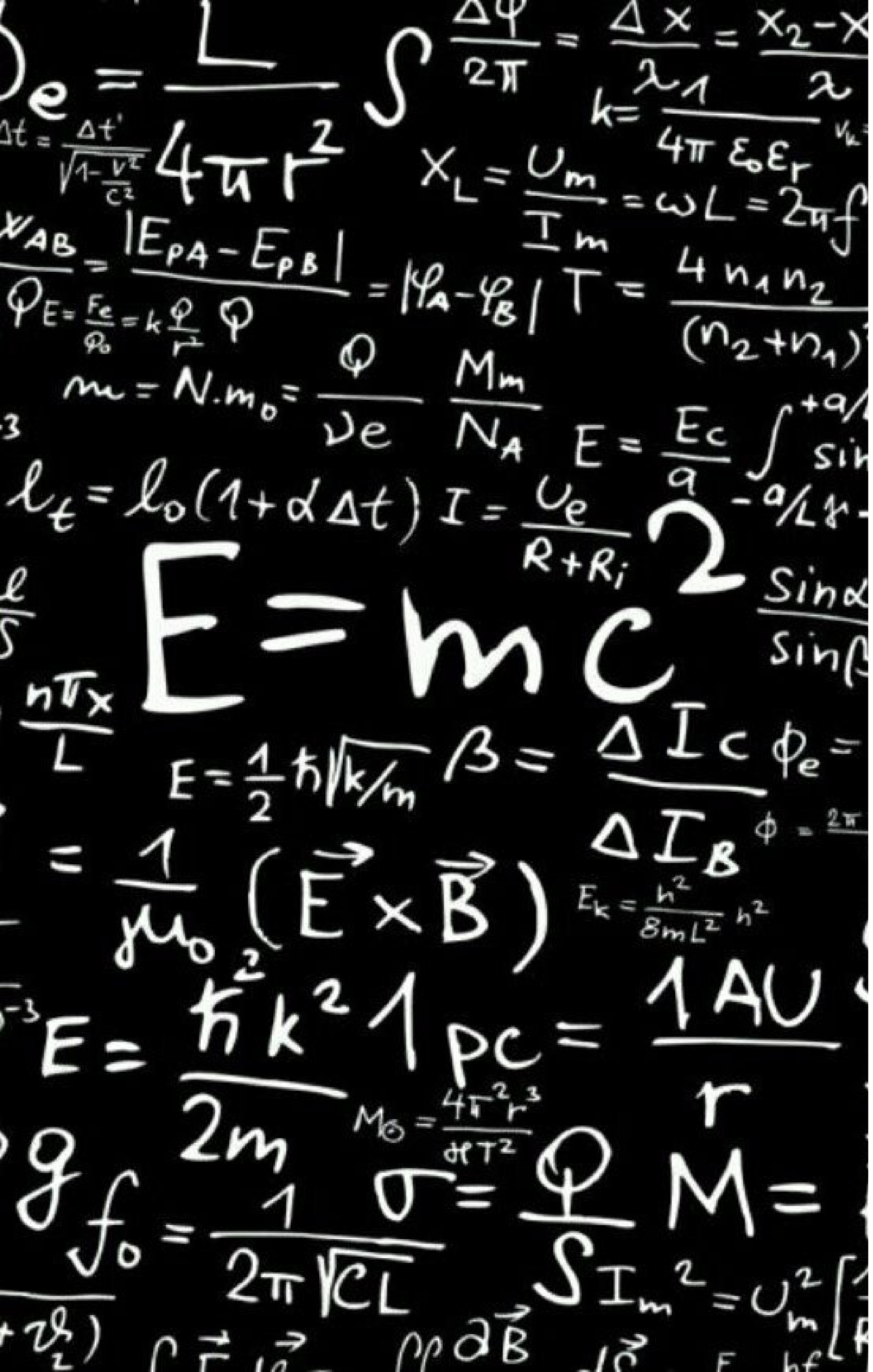
INPUT - MODEL - OUTPUT



Differences between AI and ML?



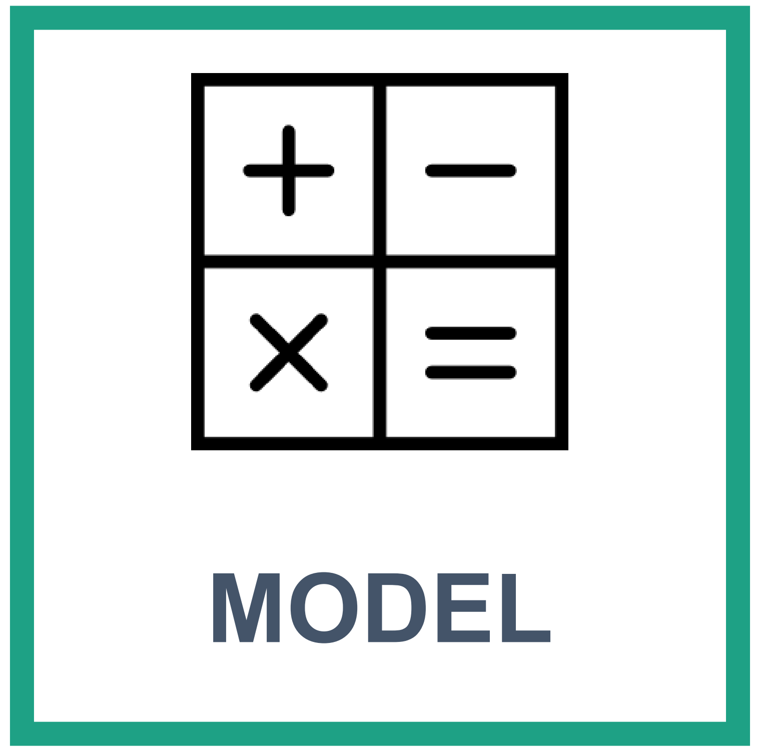
https://www.anexas.net/artificial-intelligence-online-course-in-mumbai



CLASSIFICATION AND REGRESSION

REGRESSION PROBLEM

- Property size
- Property age
- Bedrooms
- Bathrooms
- Parking size

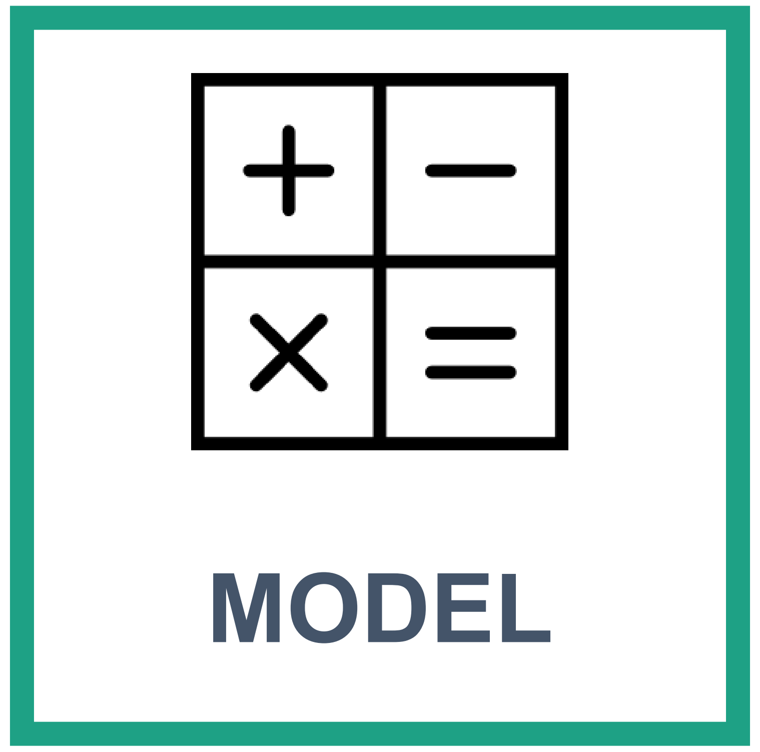


How much should we sell the property? (the answers range from 0 to 1B)

Regression problems are the type of problems where model’s answers are continuous numbers.

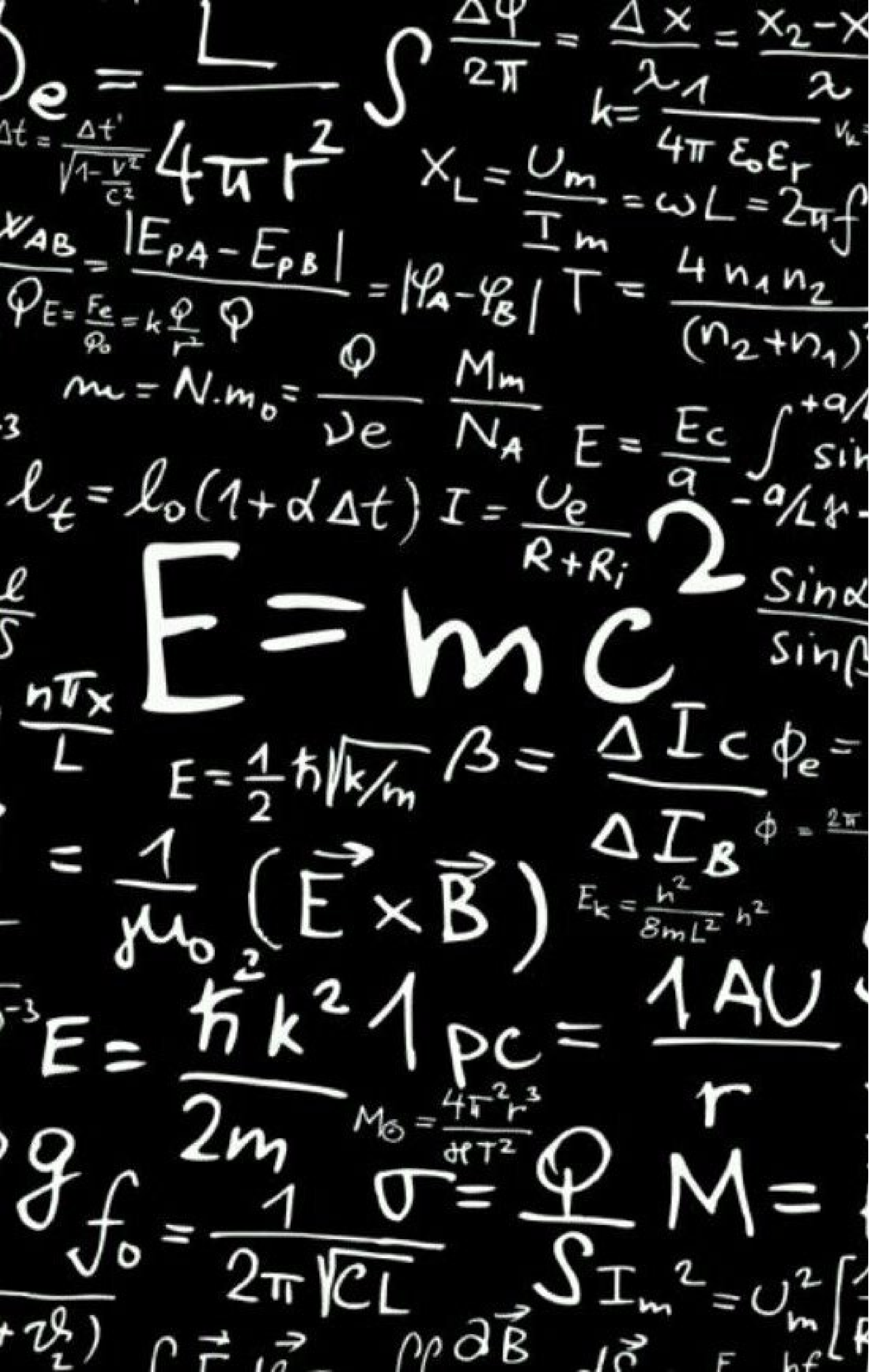
CLASSIFICATION PROBLEM

- Property size
- Property age
- Bedrooms
- Bathrooms
- Parking size



Tell me, what type of property is this? (residential or commercial)

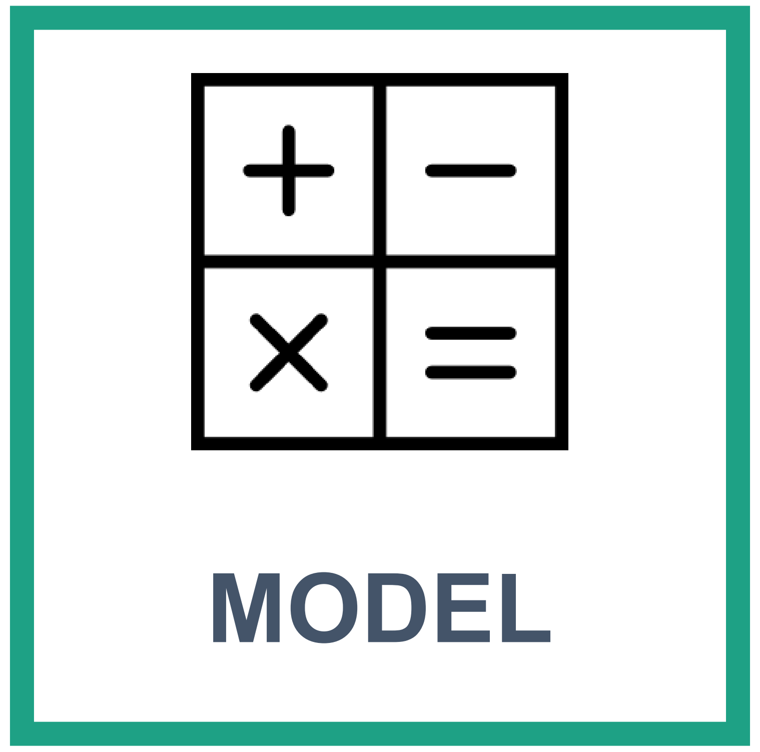
Classification problems are the type of problems where model's answers are discrete categories.



SUPERVISED AND UNSUPERVISED LEARNING

SUPERVISED LEARNING

- Property size
- Property age
- Bedrooms
- Bathrooms
- Parking size



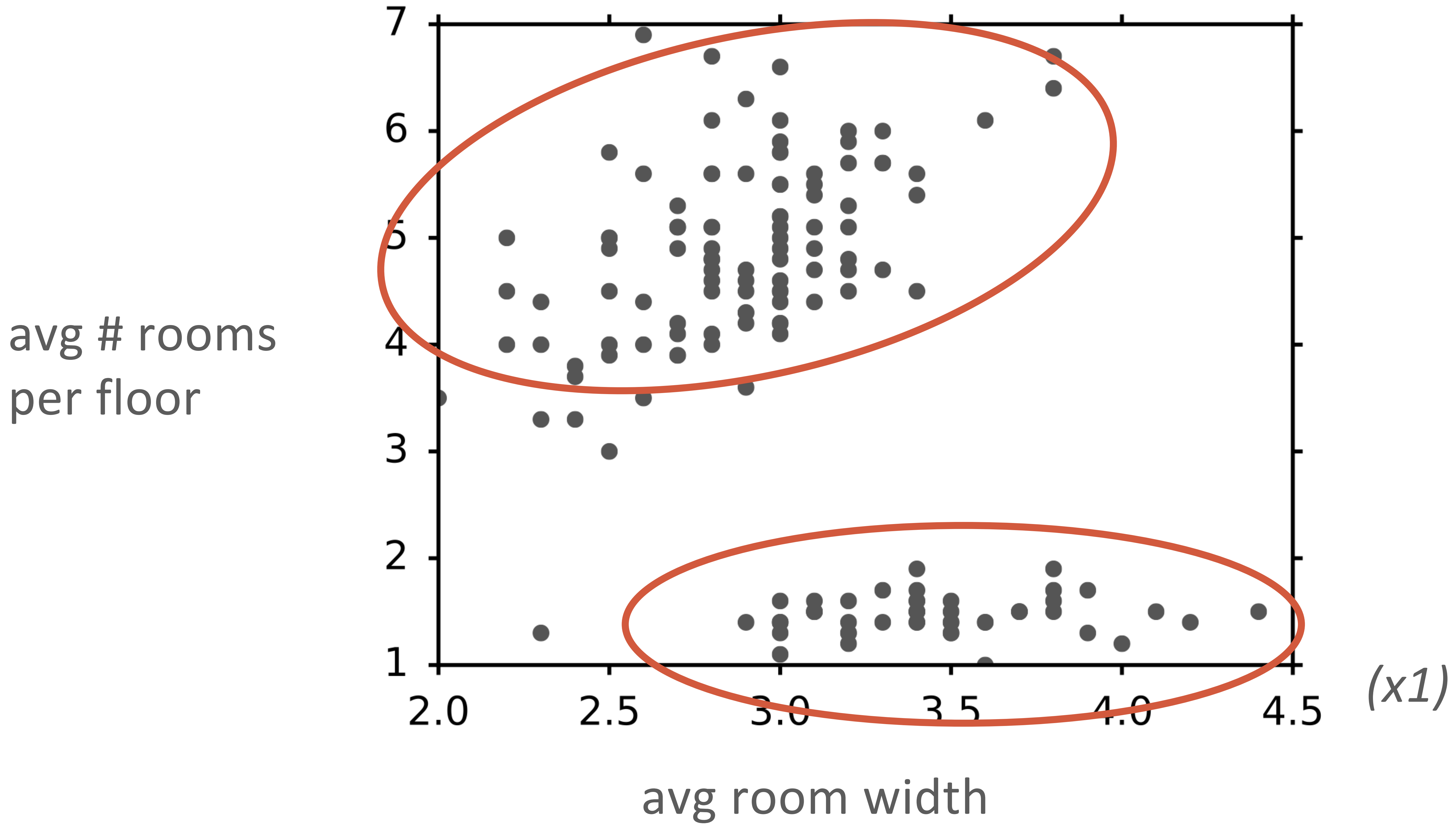
REGRESSION

How much should we sell the property? (the answers range from 0 to 1B)

CLASSIFICATION

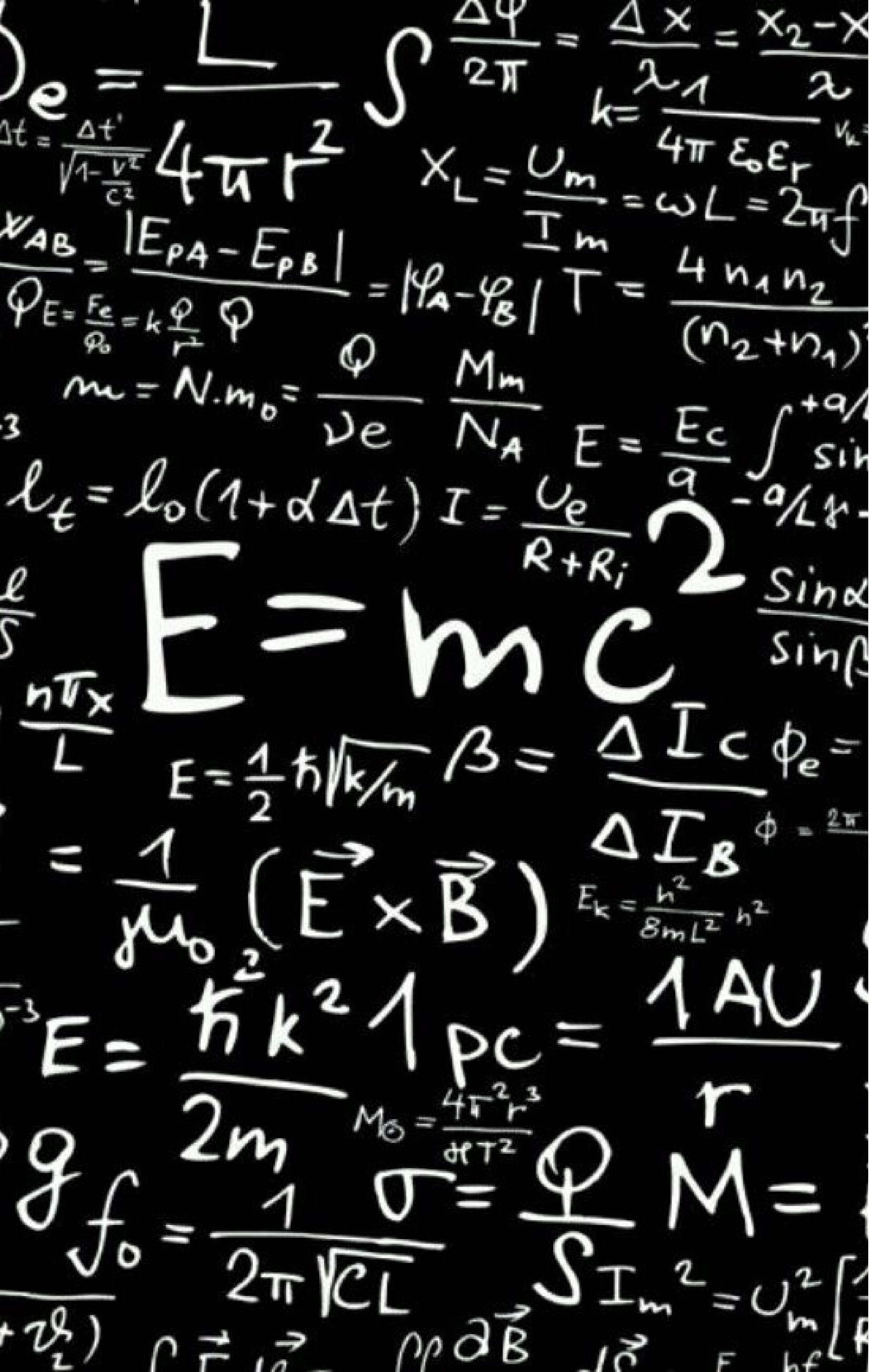
Tell me, what type of property is this? (residential or commercial)

UNSUPERVISED LEARNING



- What if you want to predict property categories, but don't know the answers in advance?
- You have to infer categories from the structure of the data.
- You are going to use unsupervised learning in this case.

Unsupervised learning problems are those problems where the answers that the model predict are not available in the training set. We infer categories from data structure.



ML PROCESS

ML Cycle

Gathering Data

Data Acquisition

- Pre-existing dataset
- Survey
- Internet

ML Cycle

Gathering Data

Data Preprocessing

Prepare the data for the model to learn

- Usually, the data is ‘dirty’. We need to ‘clean’ it.
- Missing data, bad distribution, skew data

Feature Engineering

- The gathered data may not be in the form that we want
- We need to transform some features of the dataset

ML Cycle

Gathering Data

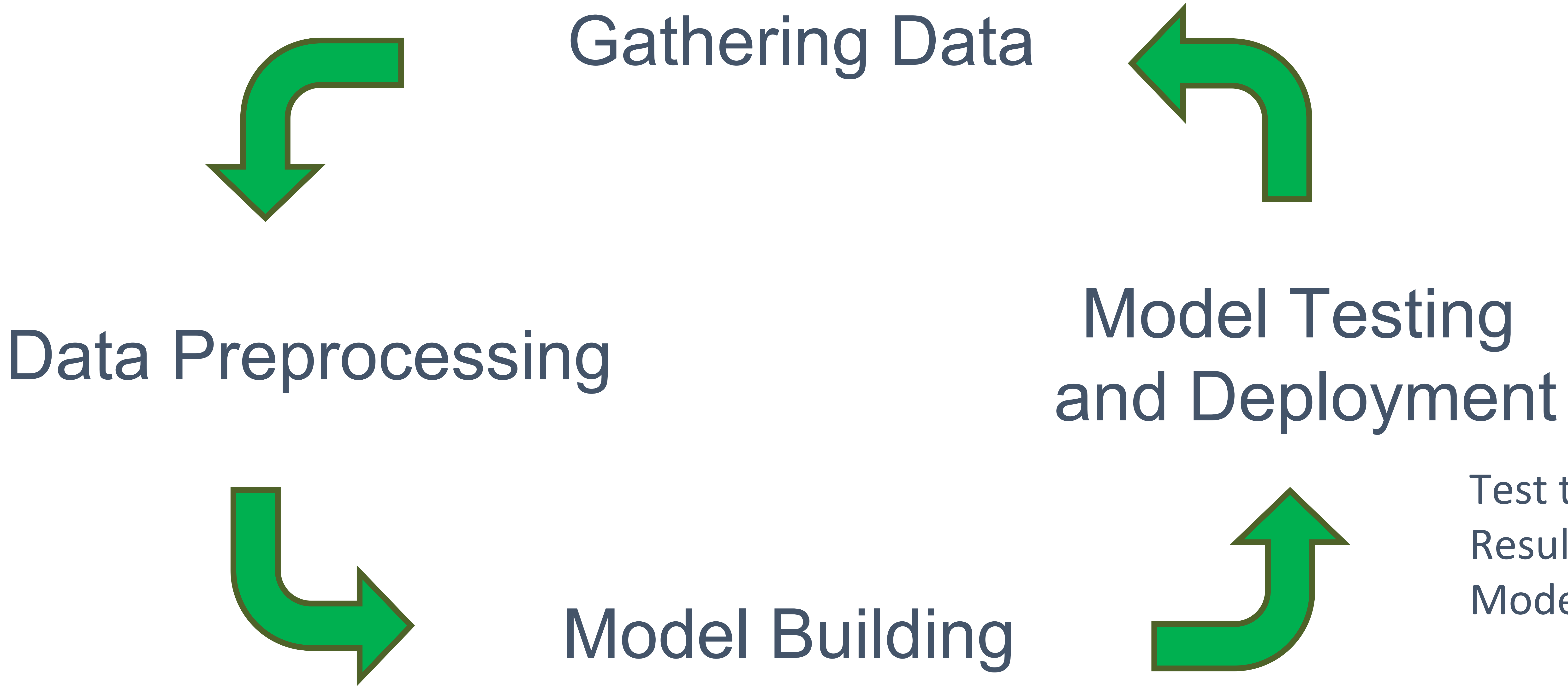
Data Preprocessing

Model Building

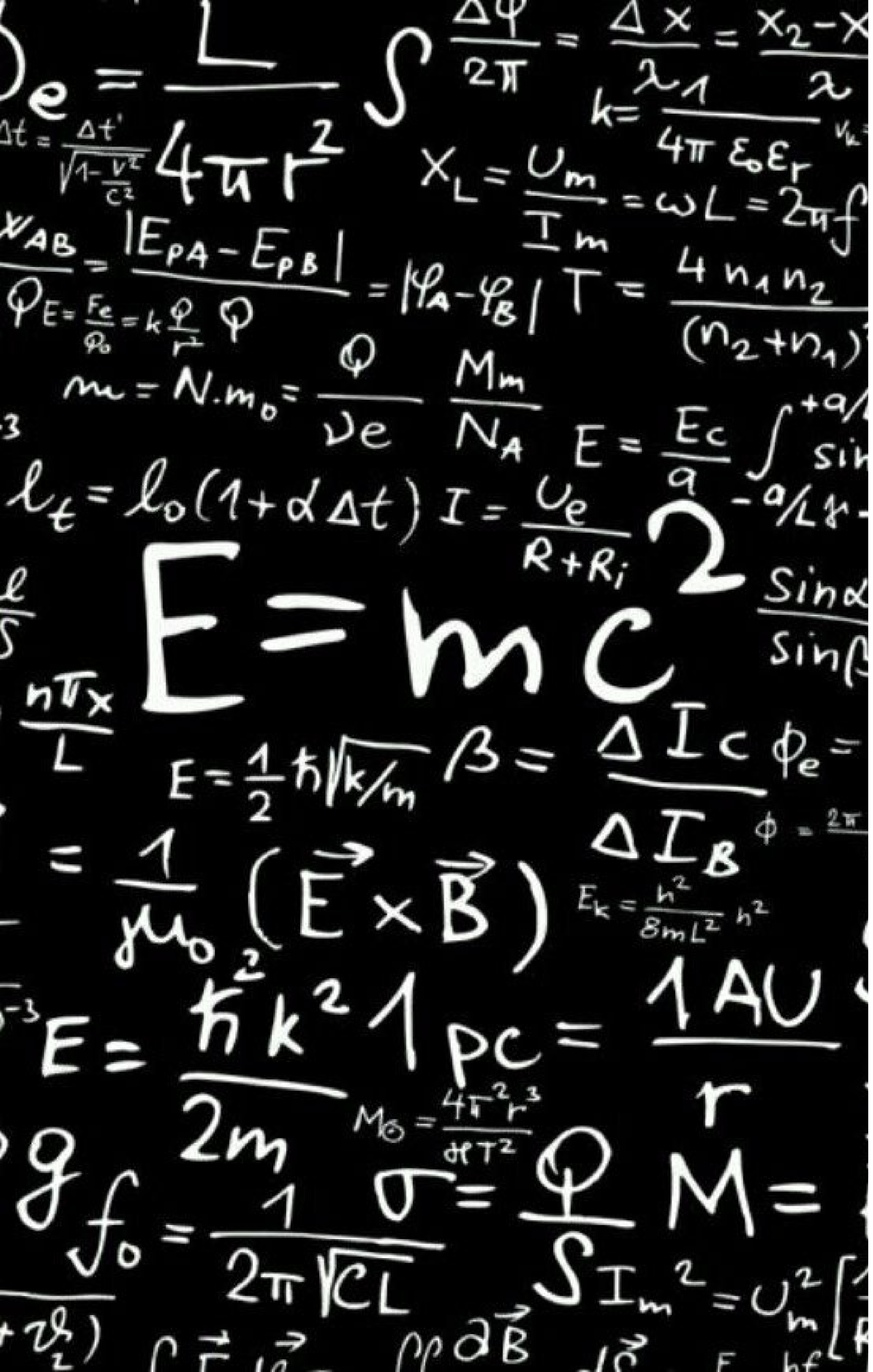
Build and Train the model

Tune the model

ML Cycle



Test the model
Result analysis
Model Deployment



UNDERSTANDING THE DATASET

Introducing Iris

- Iris is the name of a genus of flowers.
- There are 260-300 species of Iris.
- Can a machine classify the species of Iris, given a dataset?

Iris Dataset

	A	B	C	D	E
1	sepal_length	sepal_width	petal_length	petal_width	species
2	5	2	3.5	1	versicolor
3	6	2.2	4	1	versicolor
4	6	2.2	5	1.5	virginica
5	6.2	2.2	4.5	1.5	versicolor
6	4.5	2.3	1.3	0.3	setosa
7	5	2.3	3.3	1	versicolor
8	5.5	2.3	4	1.3	versicolor
9	6.3	2.3	4.4	1.3	versicolor
10	4.9	2.4	3.3	1	versicolor
11	5.5	2.4	3.8	1.1	versicolor
12	5.5	2.4	3.7	1	versicolor
13	4.9	2.5	4.5	1.7	virginica
14	5.1	2.5	3	1.1	versicolor
15	5.5	2.5	4	1.3	versicolor
16	5.6	2.5	3.9	1.1	versicolor

Attributes

	A	B	C	D	E
1	sepal_length	sepal_width	petal_length	petal_width	species
2	5	2	3.5	1	versicolor
3	6	2.2	4	1	versicolor
4	6	2.2	5	1.5	virginica
5	6.2	2.2	4.5	1.5	versicolor
6	4.5	2.3	1.3	0.3	setosa
7	5	2.3	3.3	1	versicolor
8	5.5	2.3	4	1.3	versicolor
9	6.3	2.3	4.4	1.3	versicolor
10	4.9	2.4	3.3	1	versicolor
11	5.5	2.4	3.8	1.1	versicolor
12	5.5	2.4	3.7	1	versicolor
13	4.9	2.5	4.5	1.7	virginica
14	5.1	2.5	3	1.1	versicolor
15	5.5	2.5	4	1.3	versicolor
16	5.6	2.5	3.9	1.1	versicolor

- Columns of the dataset
- It describes characteristics of each data.
- There are 2 types of attributes.
  - Numeric
  - Categorical

Numeric Attributes

	A	B	C	D	E
1	sepal_length	sepal_width	petal_length	petal_width	species
2	5	2	3.5	1	versicolor
3	6	2.2	4	1	versicolor
4	6	2.2	5	1.5	virginica
5	6.2	2.2	4.5	1.5	versicolor
6	4.5	2.3	1.3	0.3	setosa
7	5	2.3	3.3	1	versicolor
8	5.5	2.3	4	1.3	versicolor
9	6.3	2.3	4.4	1.3	versicolor
10	4.9	2.4	3.3	1	versicolor
11	5.5	2.4	3.8	1.1	versicolor
12	5.5	2.4	3.7	1	versicolor
13	4.9	2.5	4.5	1.7	virginica
14	5.1	2.5	3	1.1	versicolor
15	5.5	2.5	4	1.3	versicolor
16	5.6	2.5	3.9	1.1	versicolor

- Attributes that have numeric values.
  - aka numbers
- For example, this attribute contains information about sepal width of the flower.
- It contains numeric values of the width of the sepal.

Categorical Attributes

	A	B	C	D	E
1	sepal_length	sepal_width	petal_length	petal_width	species
2	5	2	3.5	1	versicolor
3	6	2.2	4	1	versicolor
4	6	2.2	5	1.5	virginica
5	6.2	2.2	4.5	1.5	versicolor
6	4.5	2.3	1.3	0.3	setosa
7	5	2.3	3.3	1	versicolor
8	5.5	2.3	4	1.3	versicolor
9	6.3	2.3	4.4	1.3	versicolor
10	4.9	2.4	3.3	1	versicolor
11	5.5	2.4	3.8	1.1	versicolor
12	5.5	2.4	3.7	1	versicolor
13	4.9	2.5	4.5	1.7	virginica
14	5.1	2.5	3	1.1	versicolor
15	5.5	2.5	4	1.3	versicolor
16	5.6	2.5	3.9	1.1	versicolor

- Attributes which values are categorical
  - aka discrete value
- For example, this attribute describes the species of each data point.
  - There are 3 groups (categories) of Iris: versicolor, virginica, and setosa.

Class Targets

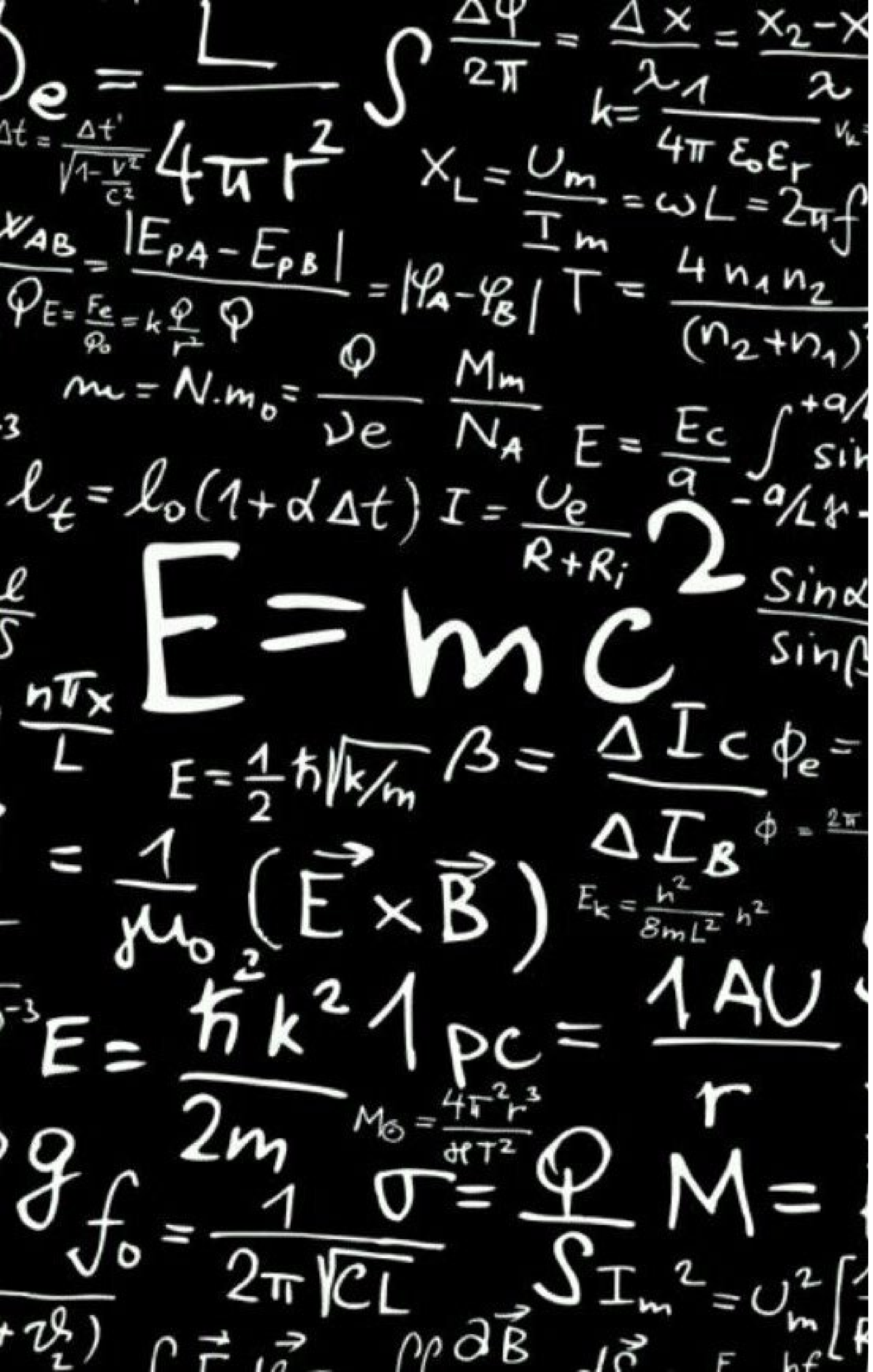
	A	B	C	D	E
1	sepal_length	sepal_width	petal_length	petal_width	species
2	5	2	3.5	1	versicolor
3	6	2.2	4	1	versicolor
4	6	2.2	5	1.5	virginica
5	6.2	2.2	4.5	1.5	versicolor
6	4.5	2.3	1.3	0.3	setosa
7	5	2.3	3.3	1	versicolor
8	5.5	2.3	4	1.3	versicolor
9	6.3	2.3	4.4	1.3	versicolor
10	4.9	2.4	3.3	1	versicolor
11	5.5	2.4	3.8	1.1	versicolor
12	5.5	2.4	3.7	1	versicolor
13	4.9	2.5	4.5	1.7	virginica
14	5.1	2.5	3	1.1	versicolor
15	5.5	2.5	4	1.3	versicolor
16	5.6	2.5	3.9	1.1	versicolor

- Attributes which classify the data.
  - It tells the class of each data point.

Instances

	A	B	C	D	E
1	sepal_length	sepal_width	petal_length	petal_width	species
2	5	2	3.5	1	versicolor
3	6	2.2	4	1	versicolor
4	6	2.2	5	1.5	virginica
5	6.2	2.2	4.5	1.5	versicolor
6	4.5	2.3	1.3	0.3	setosa
7	5	2.3	3.3	1	versicolor
8	5.5	2.3	4	1.3	versicolor
9	6.3	2.3	4.4	1.3	versicolor
10	4.9	2.4	3.3	1	versicolor
11	5.5	2.4	3.8	1.1	versicolor
12	5.5	2.4	3.7	1	versicolor
13	4.9	2.5	4.5	1.7	virginica
14	5.1	2.5	3	1.1	versicolor
15	5.5	2.5	4	1.3	versicolor
16	5.6	2.5	3.9	1.1	versicolor

- Rows of the dataset.
  - Each row describes one data point.
  - It tells the class of each data point.



MACHINE LEARNING BASICS I

HOUSE PRICE PREDICTION EXAMPLE

the most basic example

An agent has been selling 300 houses in the last years and want to be able to predict the price of a house by just knowing the size of the house.

i	Size (m^2)	Price (Mbaht)
1	50	1.4
2	128	2.6
3	24	0.8
4	78	1.2
i	…	…

the most basic example

In general,

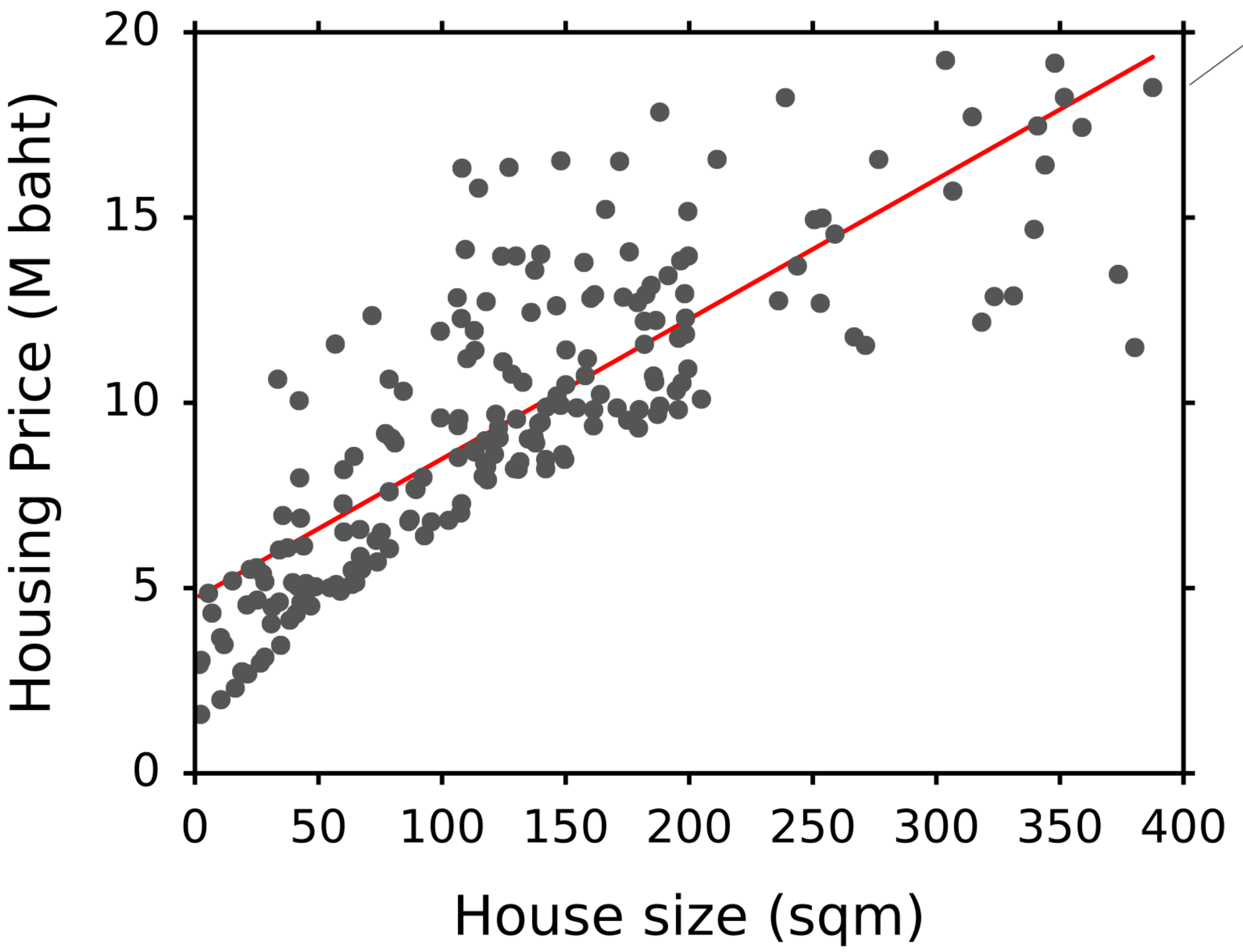
x: feature (input)

y: target (output)

i: sample index

	x		y	
i	Size (m^2)		Price (Mbaht)	
1	50	$= x_1$	1.4	$= y_1$
2	128	$= x_2$	2.6	$= y_2$
3	24	$= x_3$	0.8	$= y_3$
4	78	$= x_4$	1.2	$= y_4$
i	…	$= x_i$	…	$= y_i$

what is a model



$$h(x) = \underbrace{\theta_0}_{} + \underbrace{\theta_1}_{} x$$

parameters

of the model

$$\theta_0 = 4.717$$

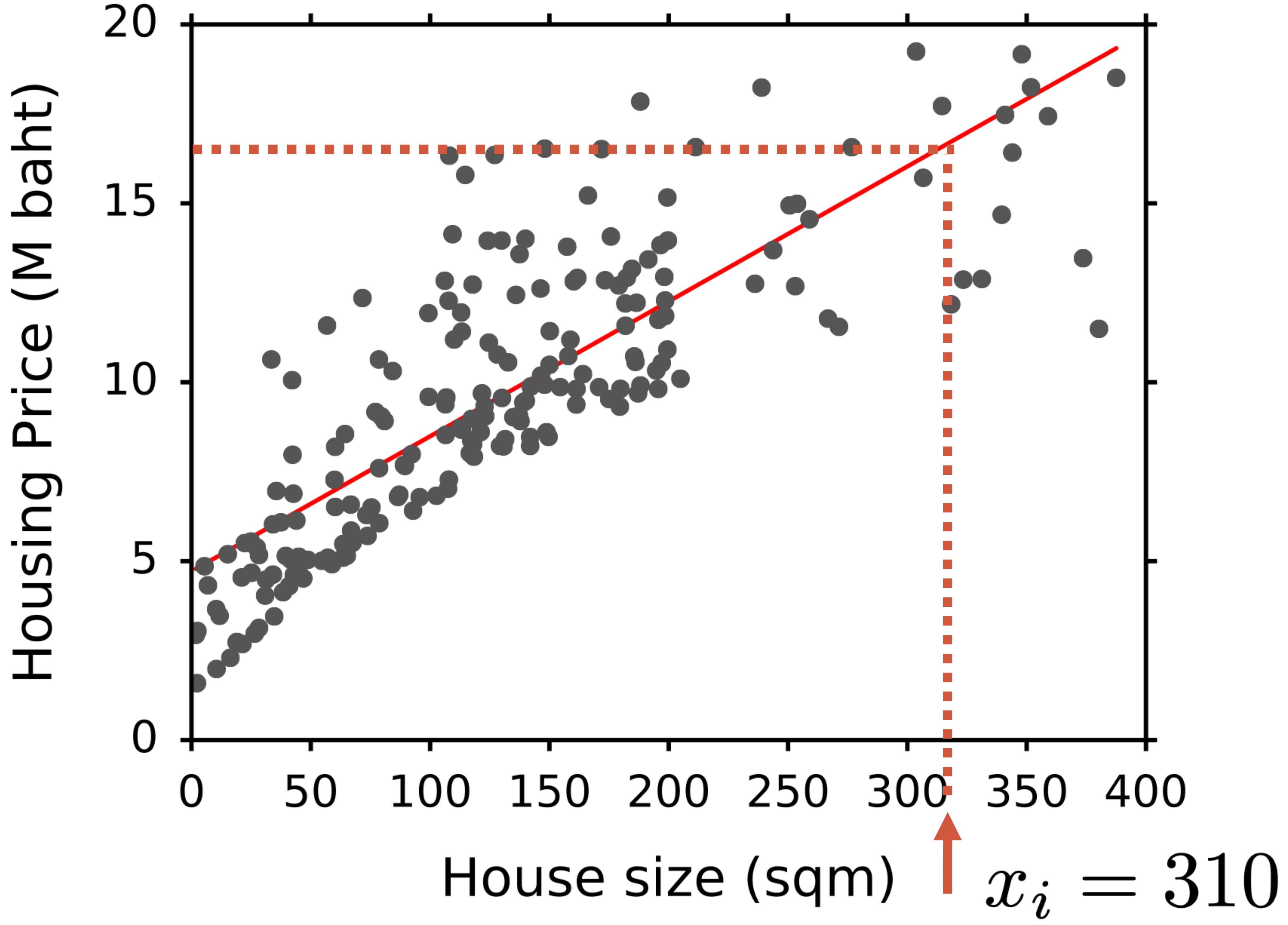
$$\theta_1 = 0.038$$

A model is a function that takes an input and yields the values we want to predict.

what is a model

After we have a model, we can predict y value from any x value

Model: $h(x_i) = 4.717 + 0.038 * (x_i)$



Notice that:

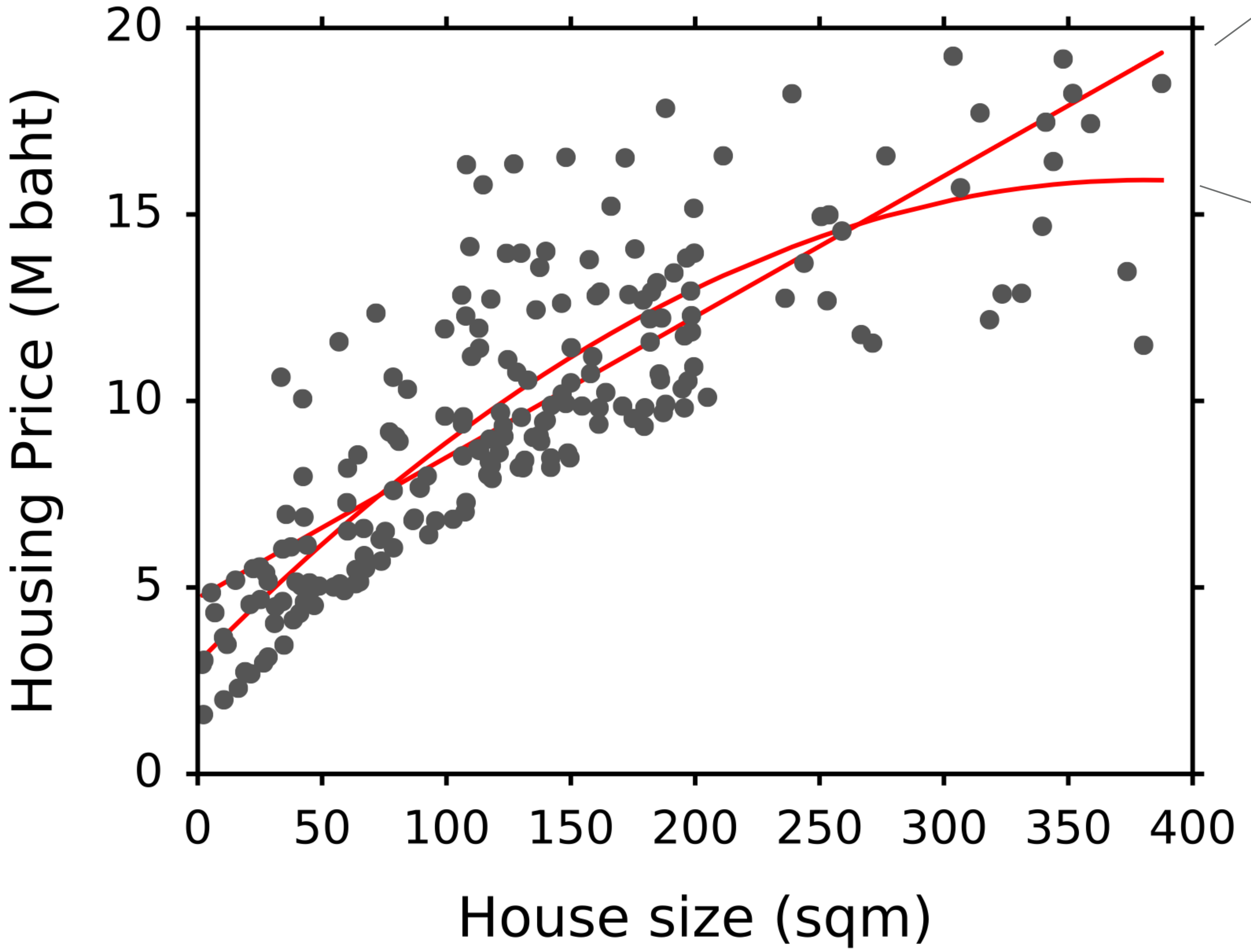
h(x) and y are not the same

h(x): value we predict

y: the actual value

what is a model

For different models you have different functions, with perhaps different number of parameters. Different models can be fitted to the same data set.



$$h(x) = \theta_0 + \theta_1 x$$

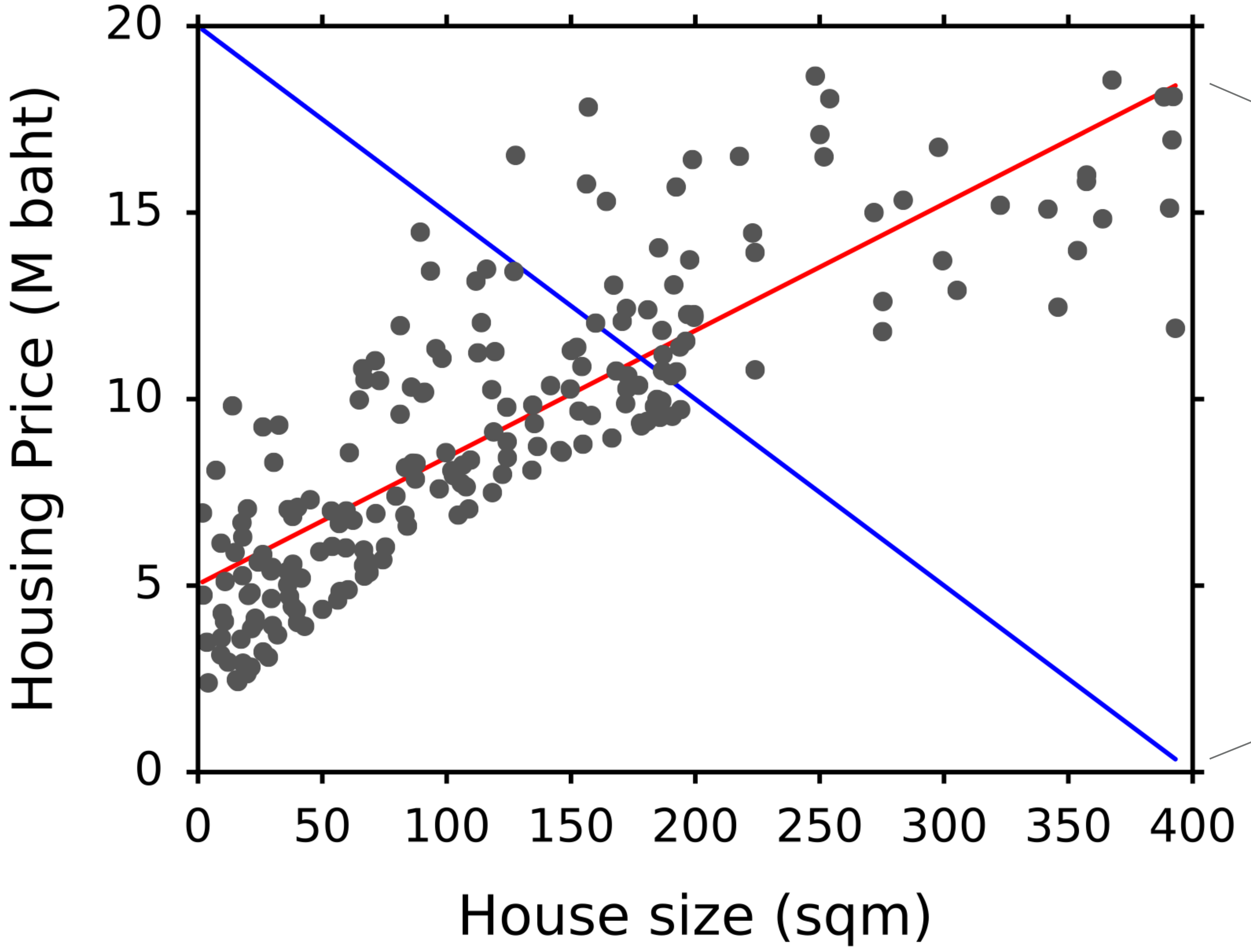
$$h(x) = \theta_0 + \theta_1 x + \theta_2 x^2$$

Model is a function of both features (x) and parameter (theta).

$$h(x; \theta)$$

good and bad models

Model: $h(x) = \theta_0 + \theta_1 x$



$\theta_0 = 4.717$

$\theta_1 = 0.038$

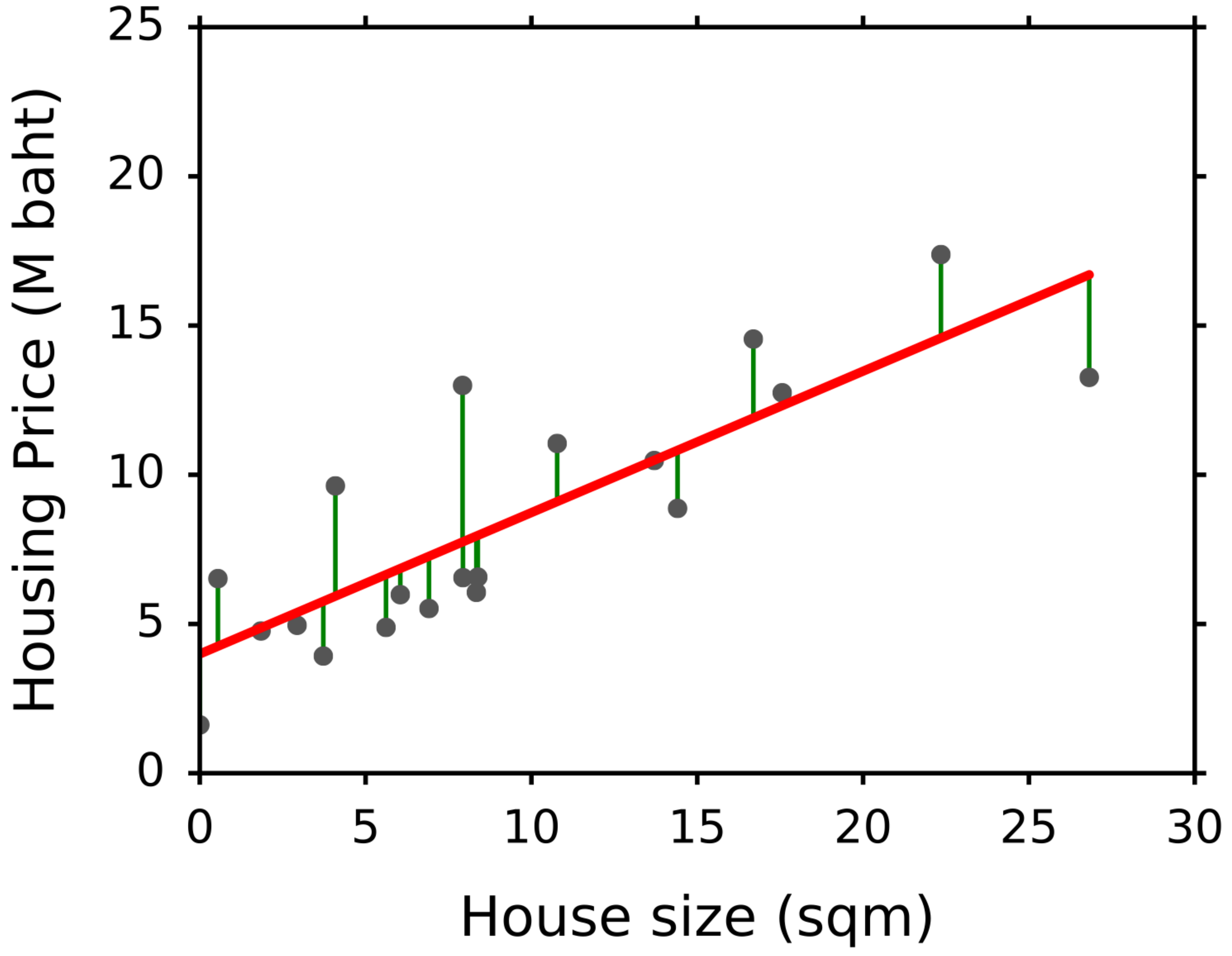
$\theta_0 = 20$

$\theta_1 = -0.05$

cost function

Cost function (or loss function) is a measure of whether a model is a good fit to the data.

For example, a famous cost function is called 'squared error' function that takes the difference between what you predict and the actual data value and square it.



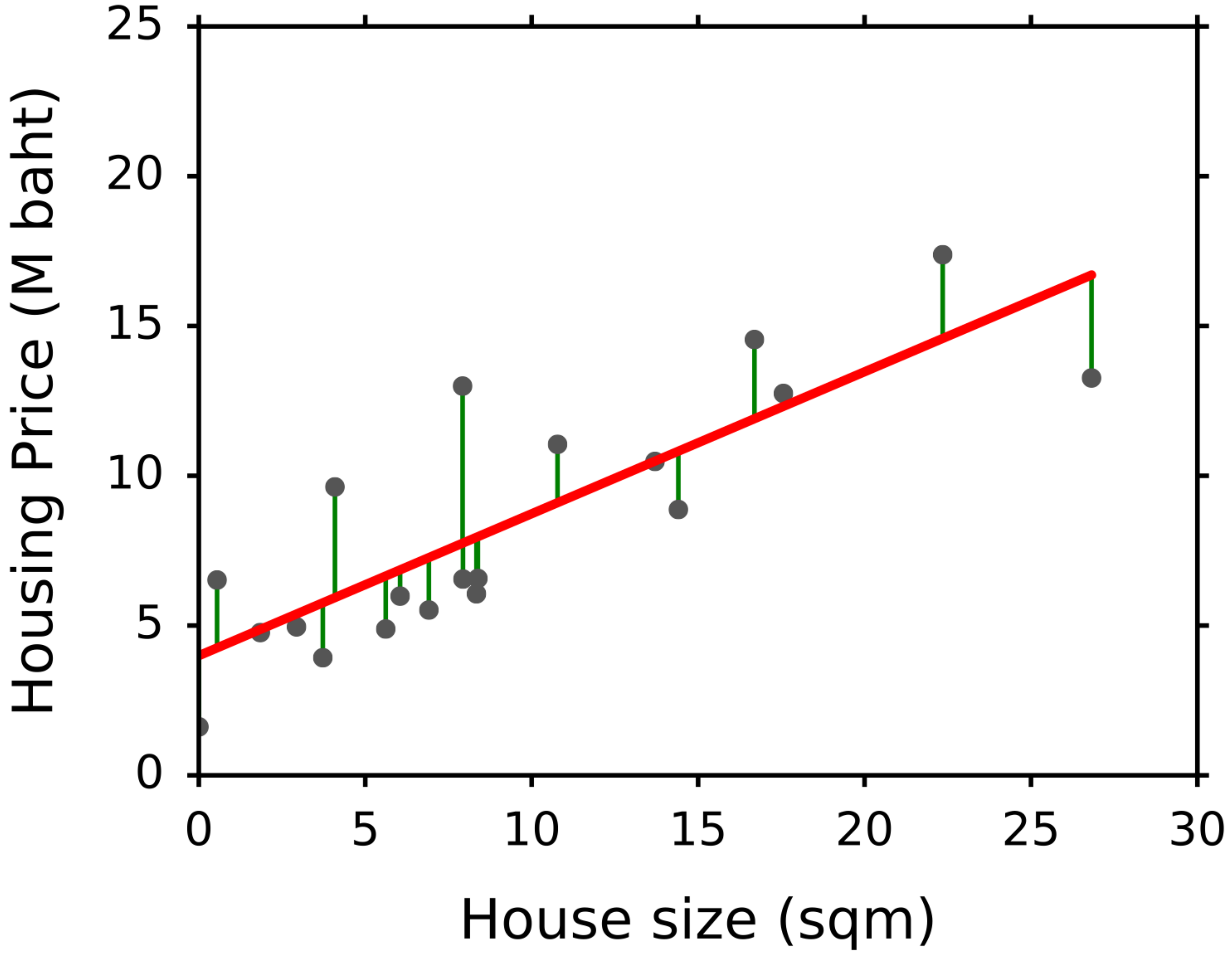
$$\sum_i (h(x_i) - y_i)^2$$

lowering cost function

Model: $h(x) = \theta_0 + \theta_1 x$

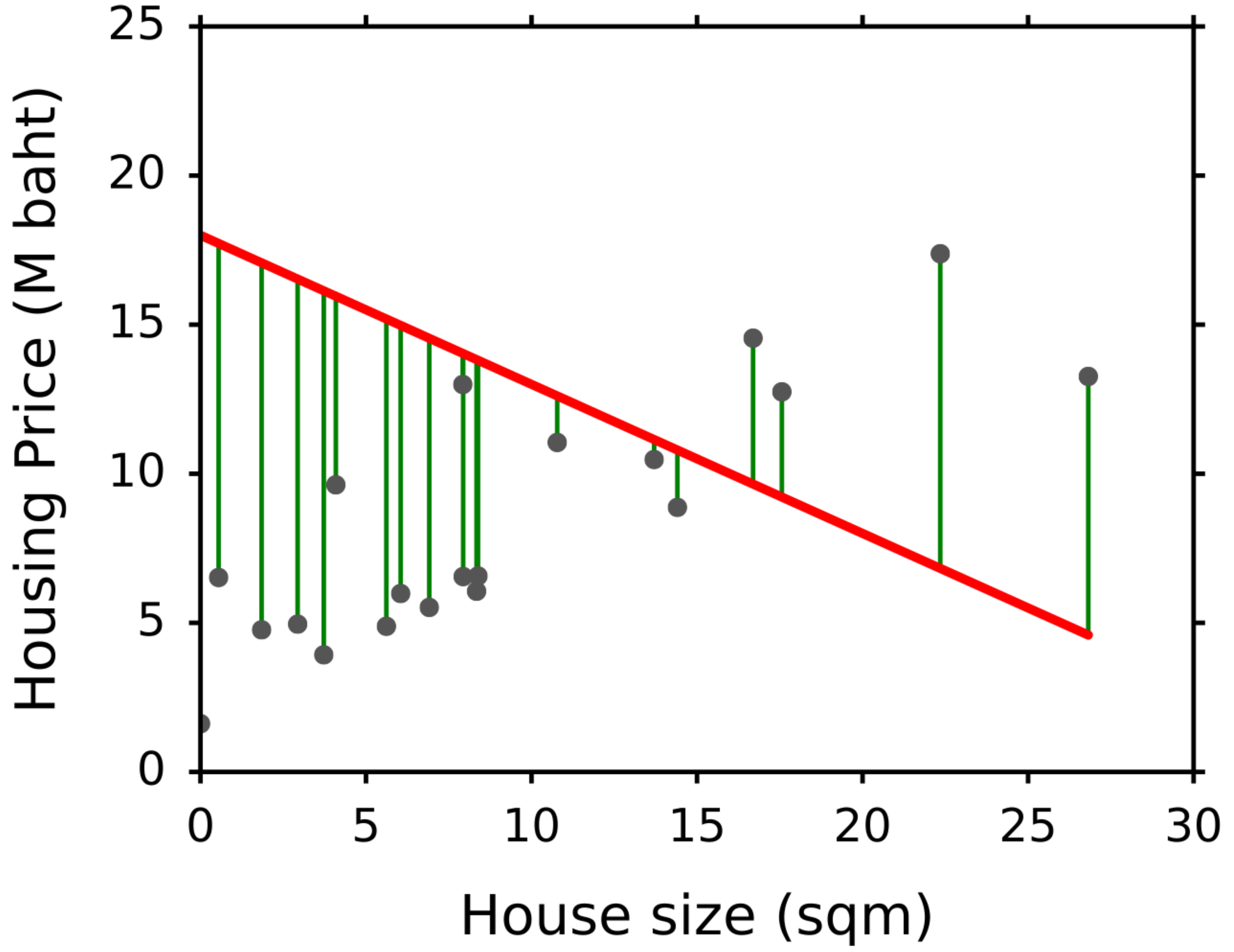
$\theta_0 = 4.717$

$\theta_1 = 0.038$



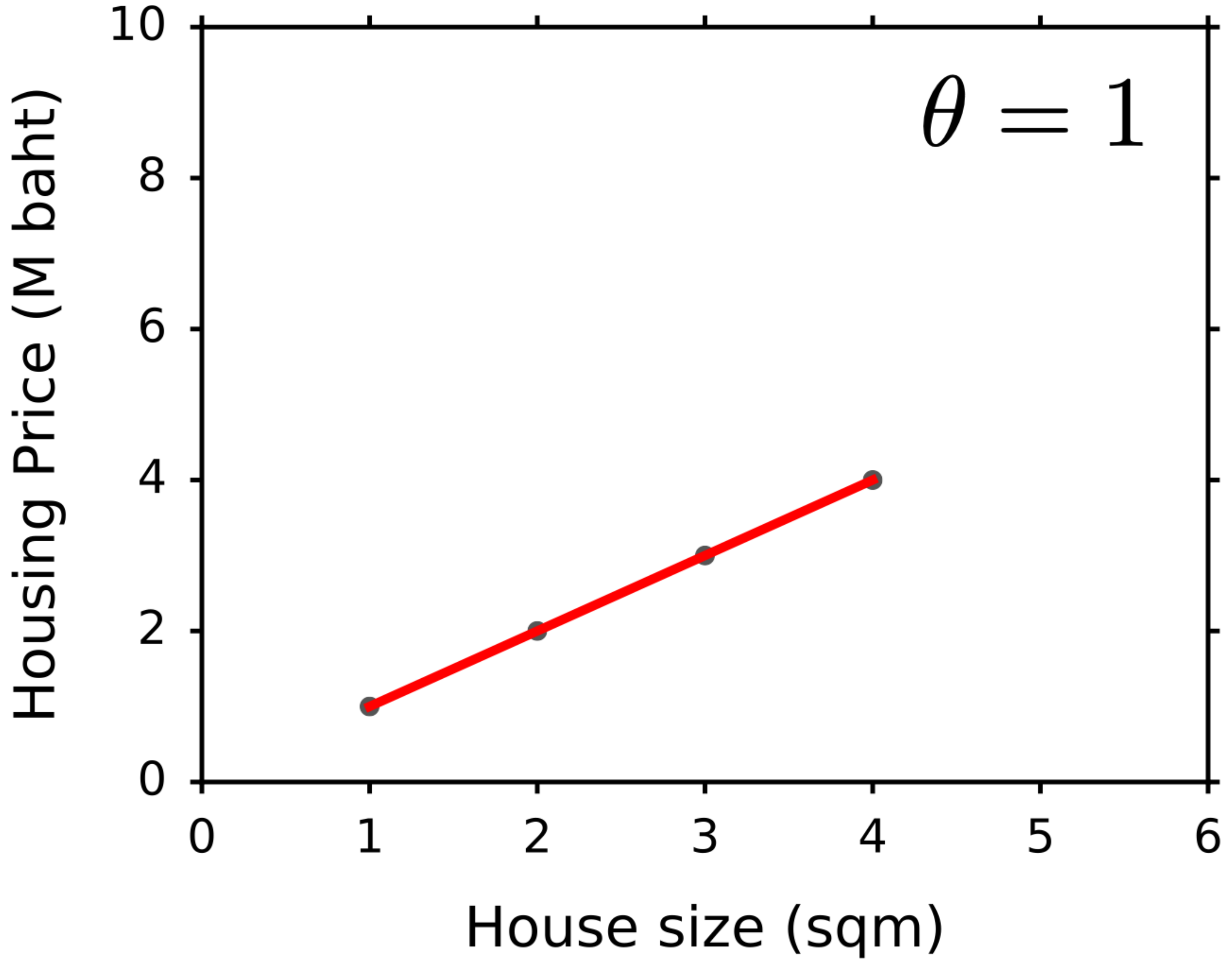
$\theta_0 = 20$

$\theta_1 = -0.05$

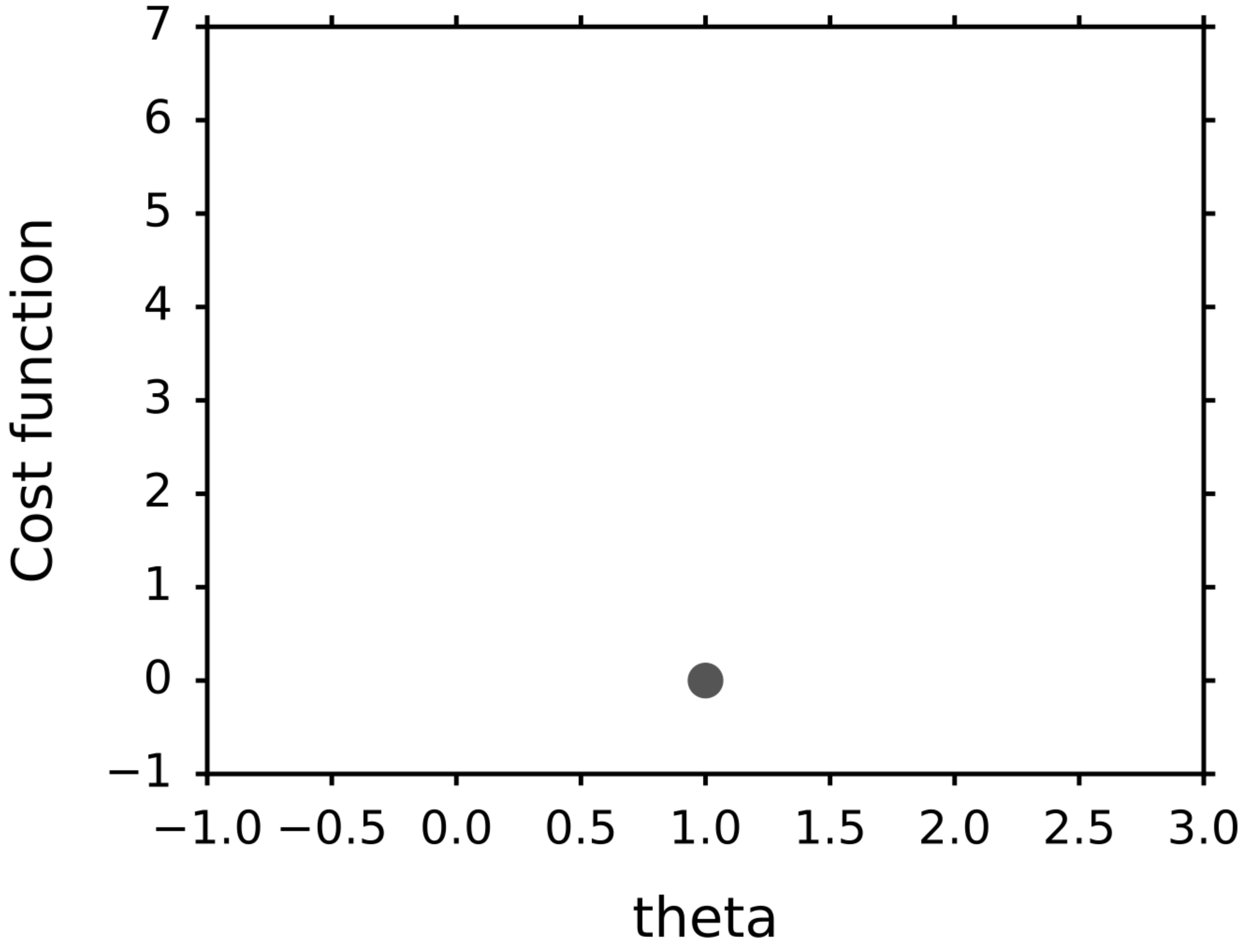


minimizing cost function

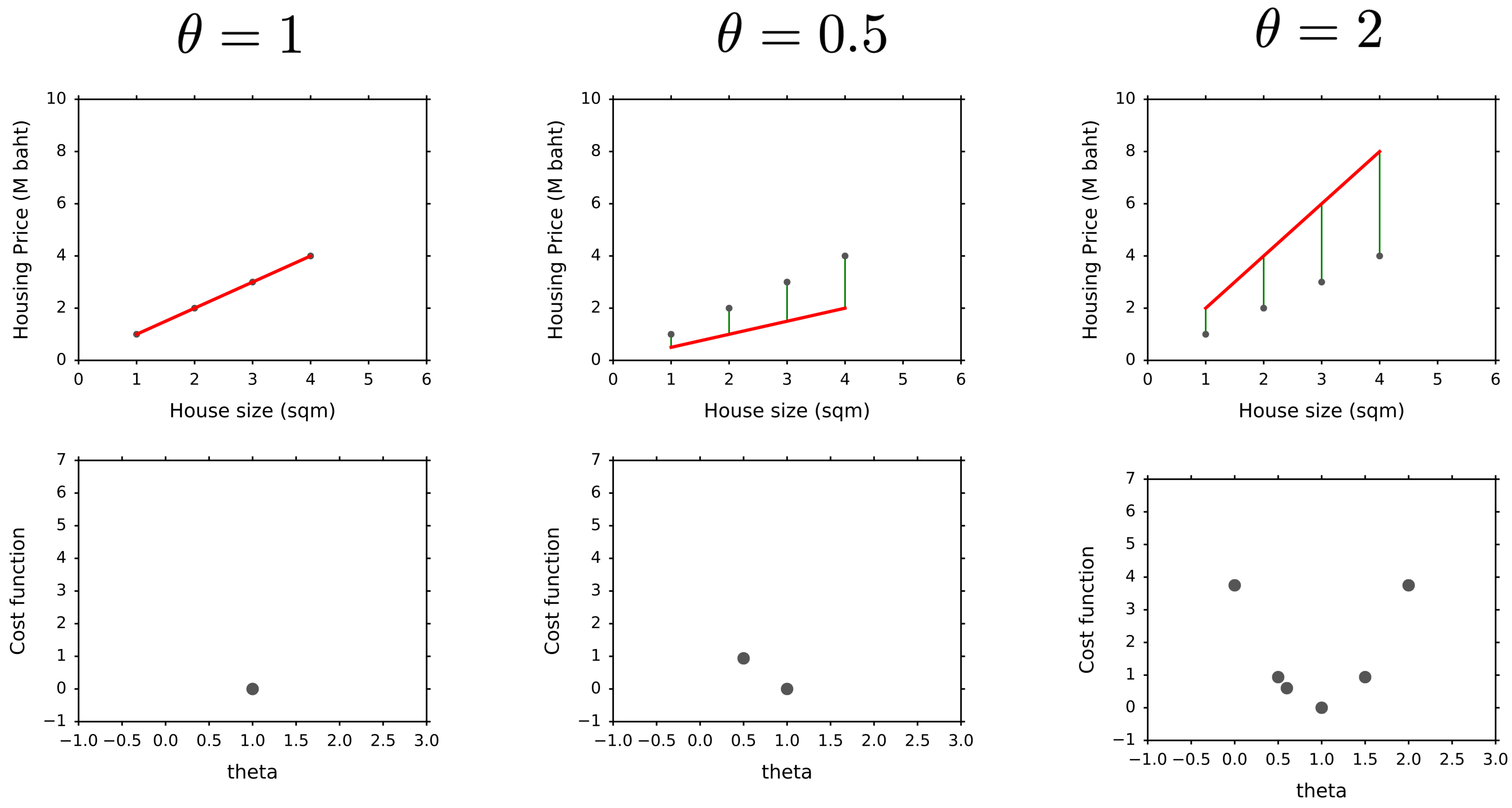
Model: $h(x) = \theta x$



$Cost(\theta)$

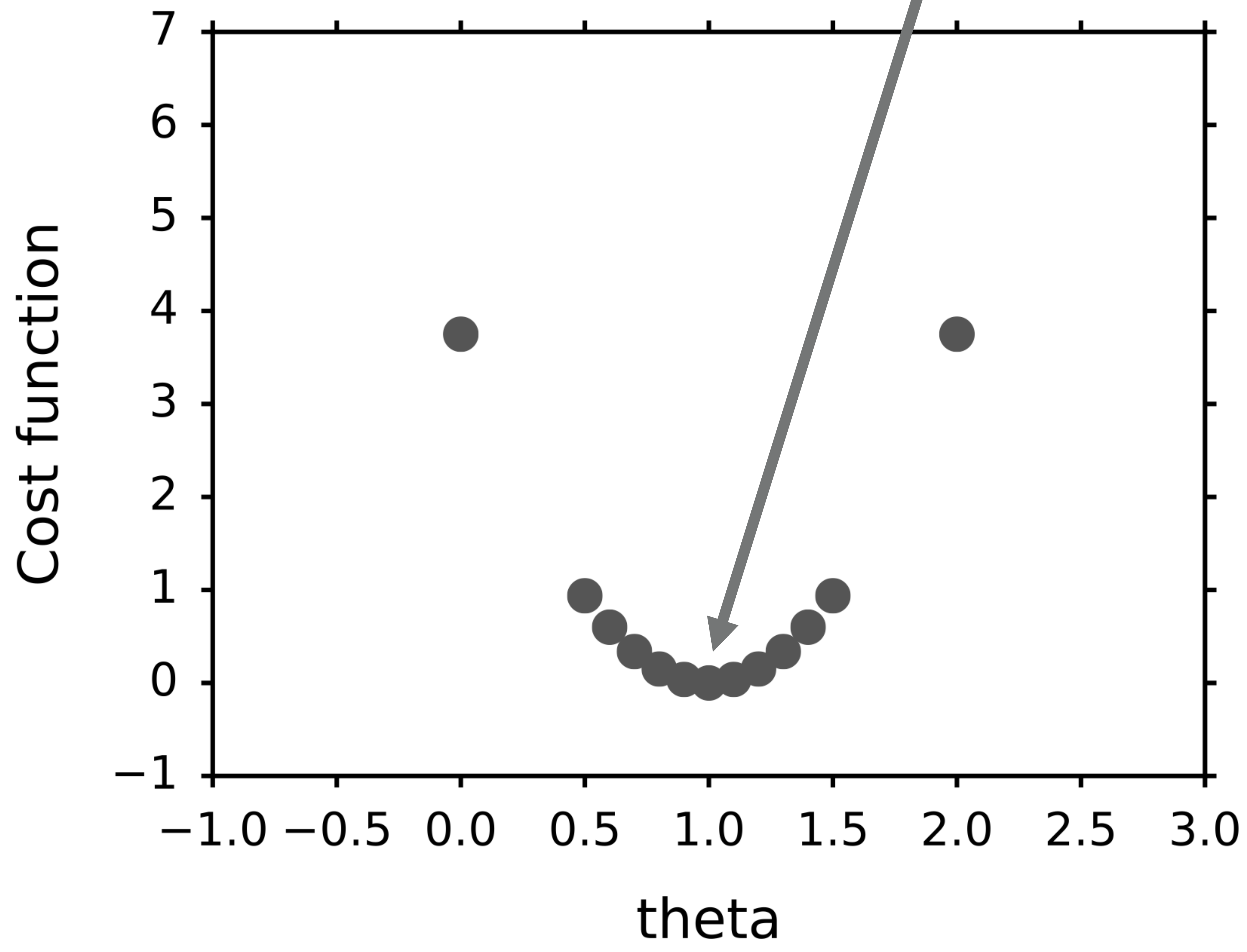


minimizing cost function



minimizing cost function

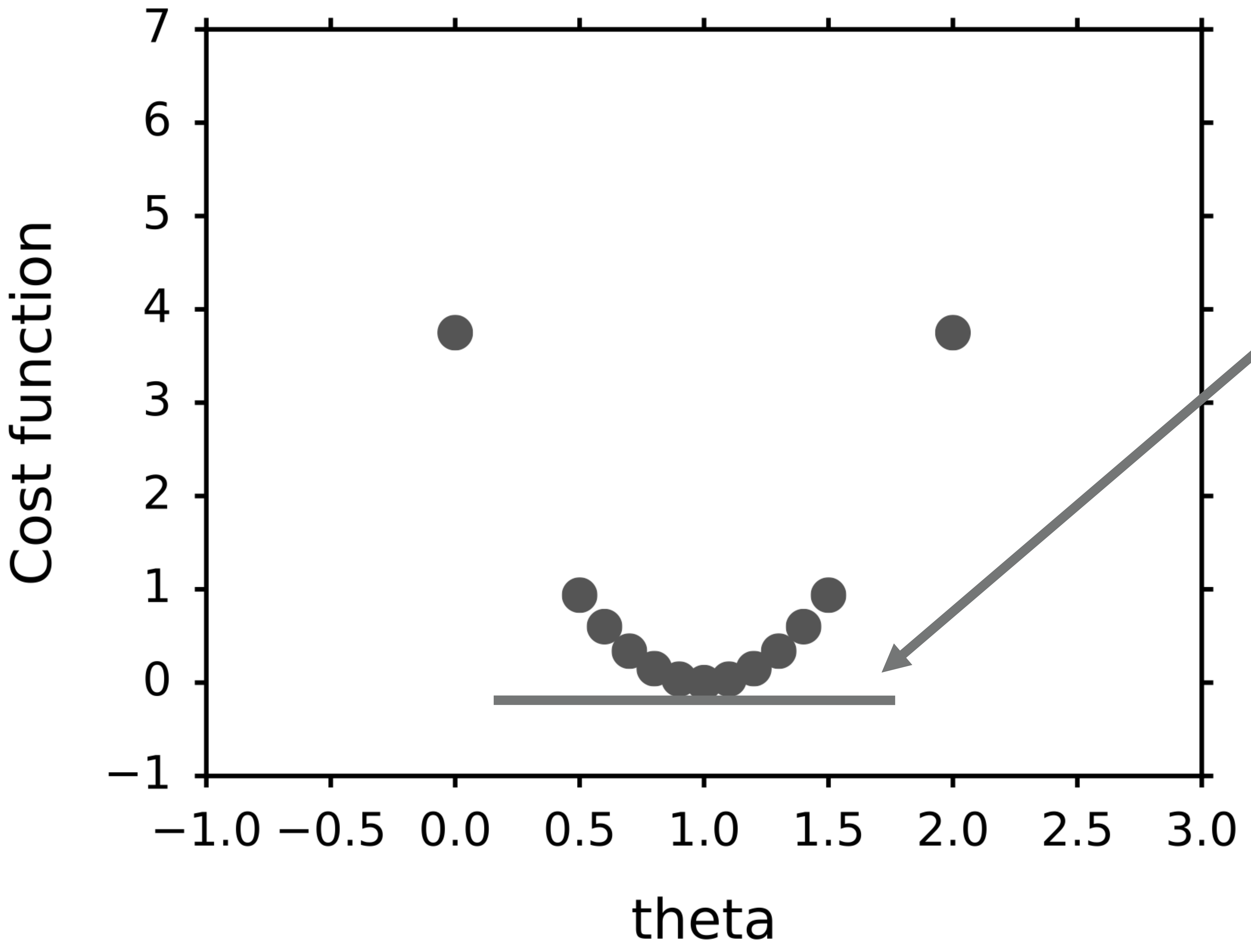
Best-fitted model has lowest cost function



Training machine learning models = minimizing cost function

Often accomplished by using ‘optimization algorithms’

minimizing cost function



$$\frac{\partial Cost}{\partial \theta} = 0$$

Some optimization algorithm is so simple,

you just solve an equation and done.

Machine learning basics

- To train the model, we collect samples, which are data points that the model need to learn.
- In sample data, we have targets (the quantity we want to predict) and features (the data we use to predict targets).
- A model is a set of equations that map features to predictions.
- All models have parameters which we can adjust to fit the dataset.

Machine learning basics

- Model's predictions (often called h(x)) are not the same as targets (y).
- Cost function is an equation that measures the difference between h(x) and y.
- To fit the model to the data, we adjust parameters in a way that minimize cost function.