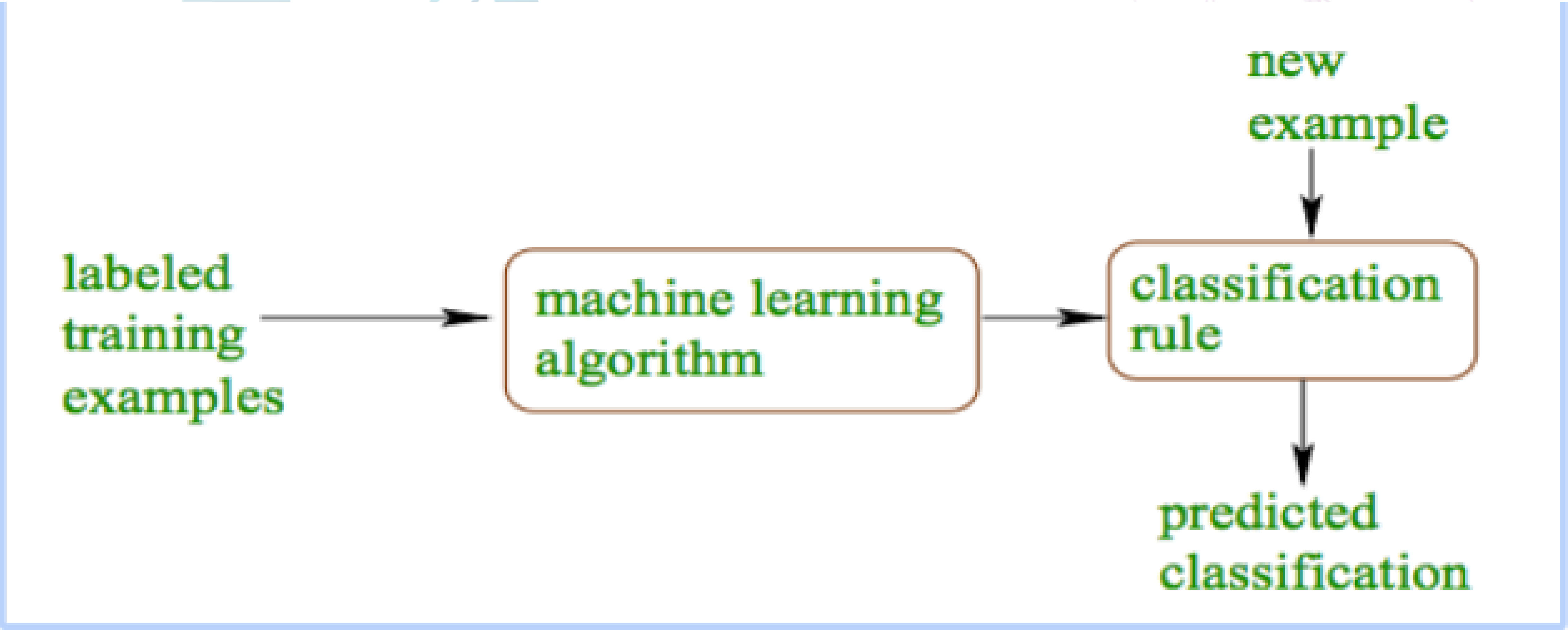
Logistic Regression

Machine Learning - Day 2

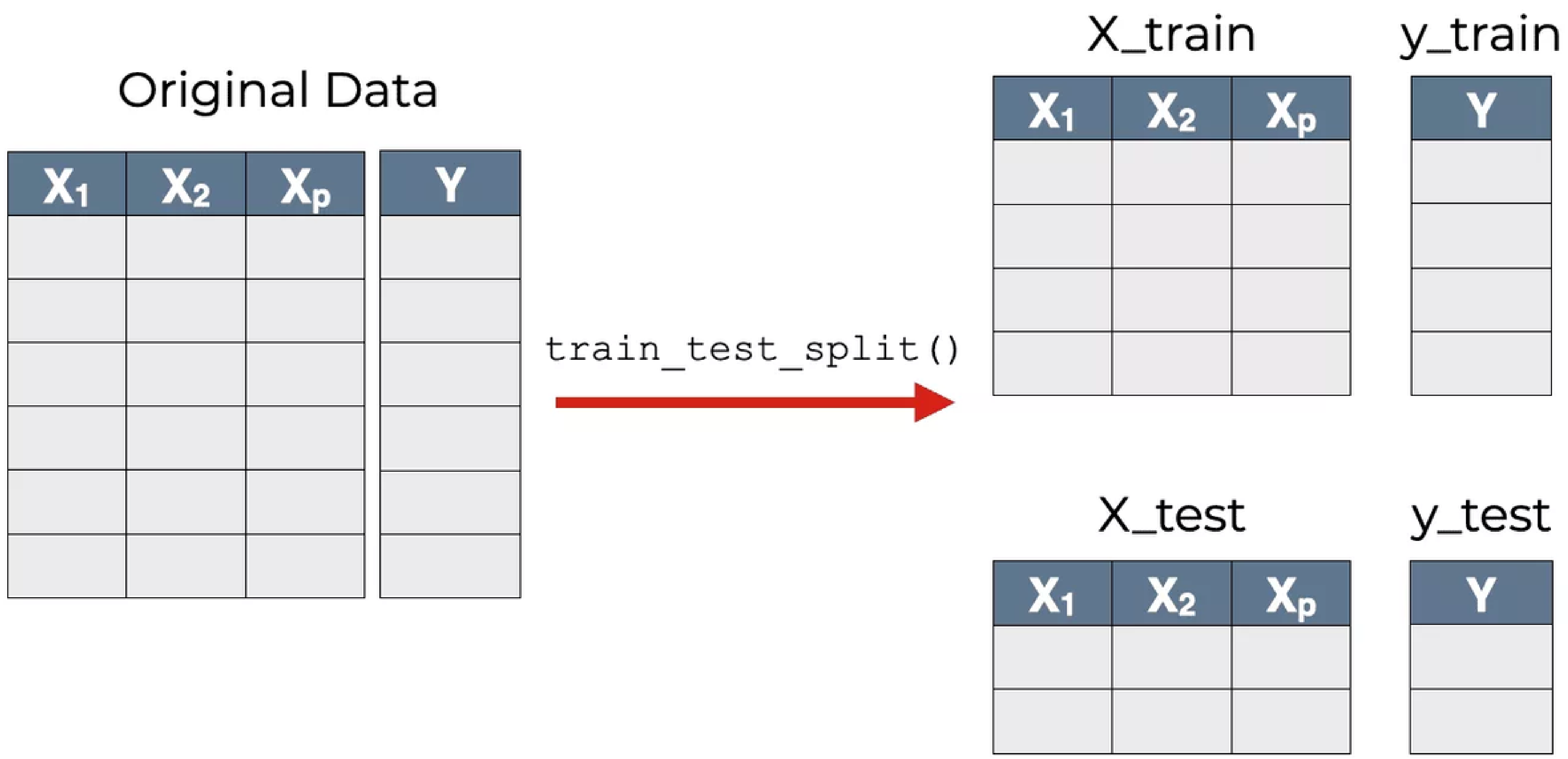
Review - Day 2

- Machine Learning
- Supervised vs Unsupervised Learning
- Regression and Classification
- Linear Regression
- Train test split
- Loss function Curves

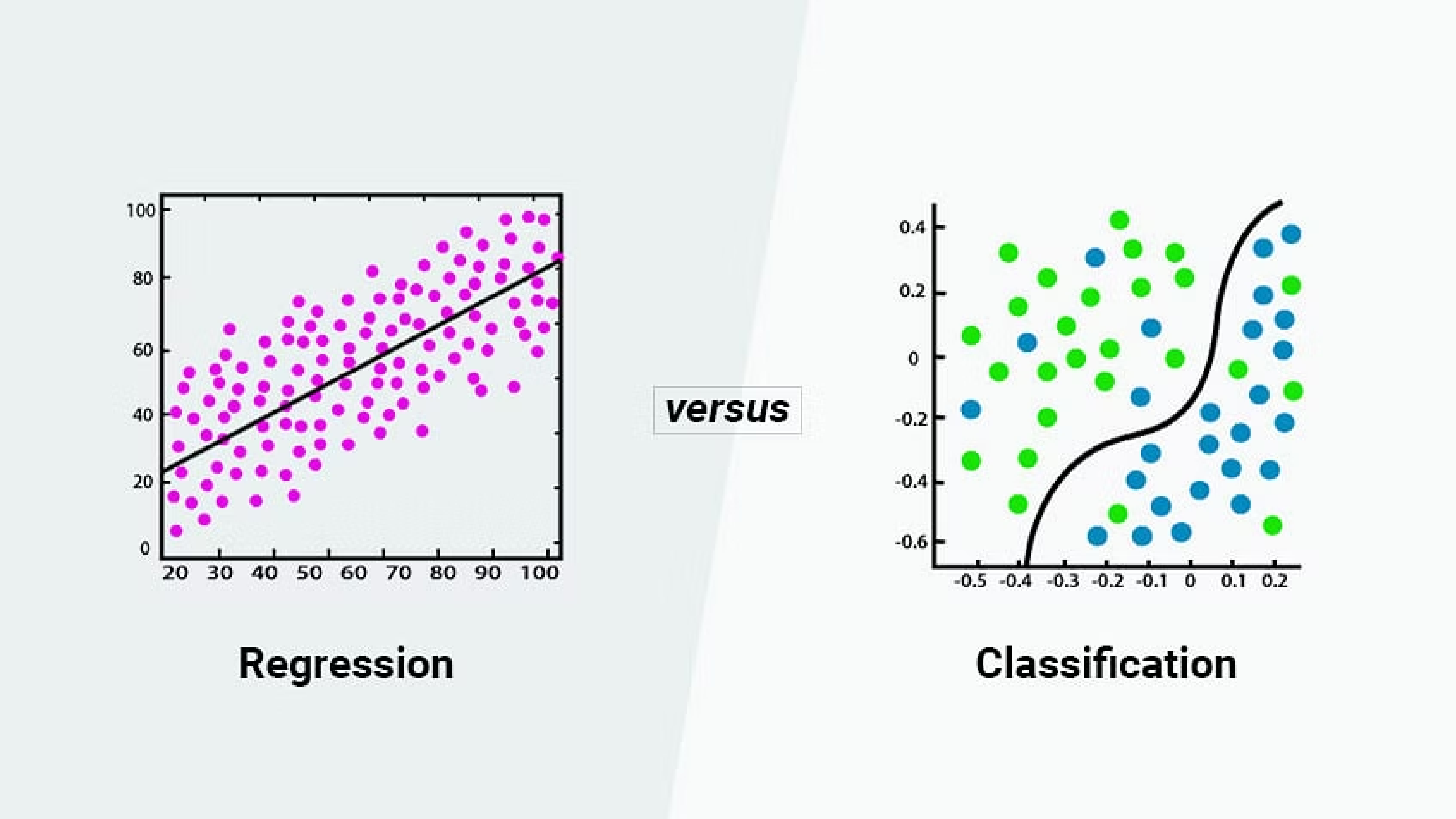


General Steps:

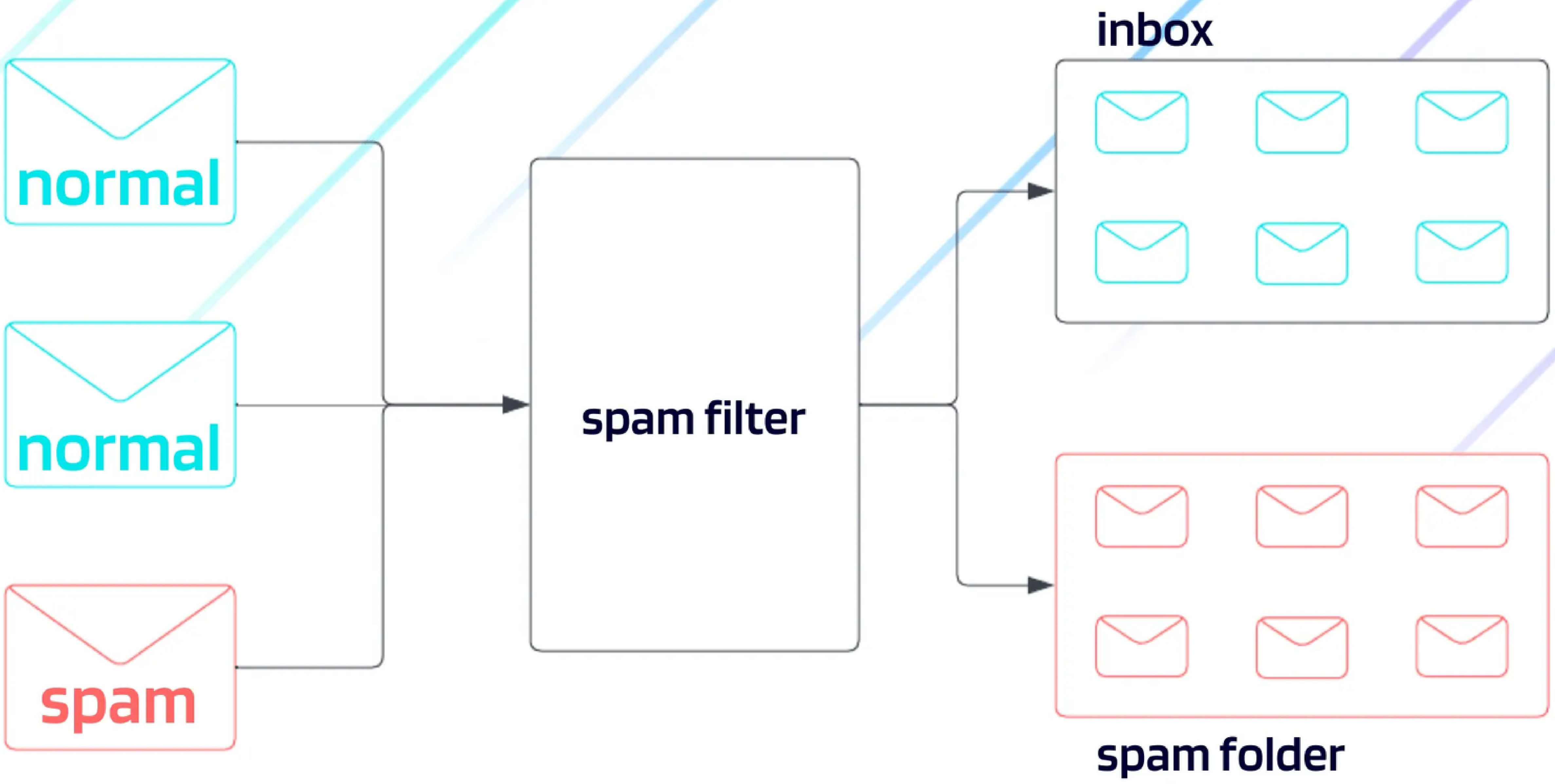
- Split data into "training" and "test" sets.
- Use regression/classification results from "training" set to predict"test" set
- Compare "Predicted Y" to "Actual Y"



Logistic Regression is used when the dependent variable(target) is categorical.



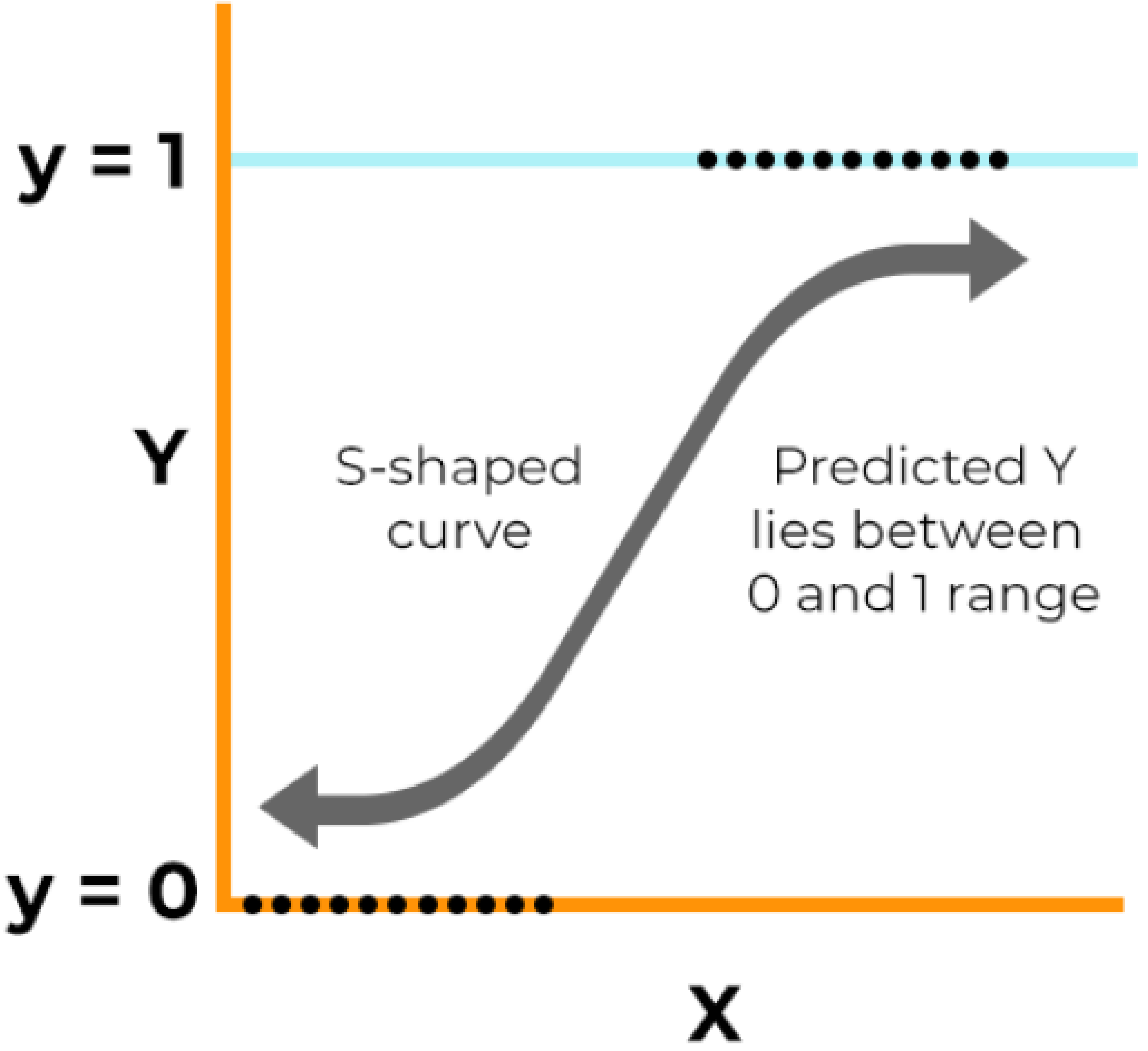
Spam vs Ham (Non-spam)



Other Examples

- Whether a person will pass in the exam or not
- Whether a person is suffering from a disease
- Whether a person likes a movie or not
- Whether a person will buy a stuff or not
- Whether a tumor is malignant or not

Logistic Regression Curve



1. Dataset Preparation

Let's say we have a dataset

Study hours (X)	Exam result (y)
2	0
3	0
4	0
5	1
6	1

2. Model Training

We fit a logistic regression model to the dataset to predict the probability of passing exam based on number of study hours.

Logistic regression model:
z=a1+a2*x

Our goal: estimate values of a1 and a2 so that data is fitted best.

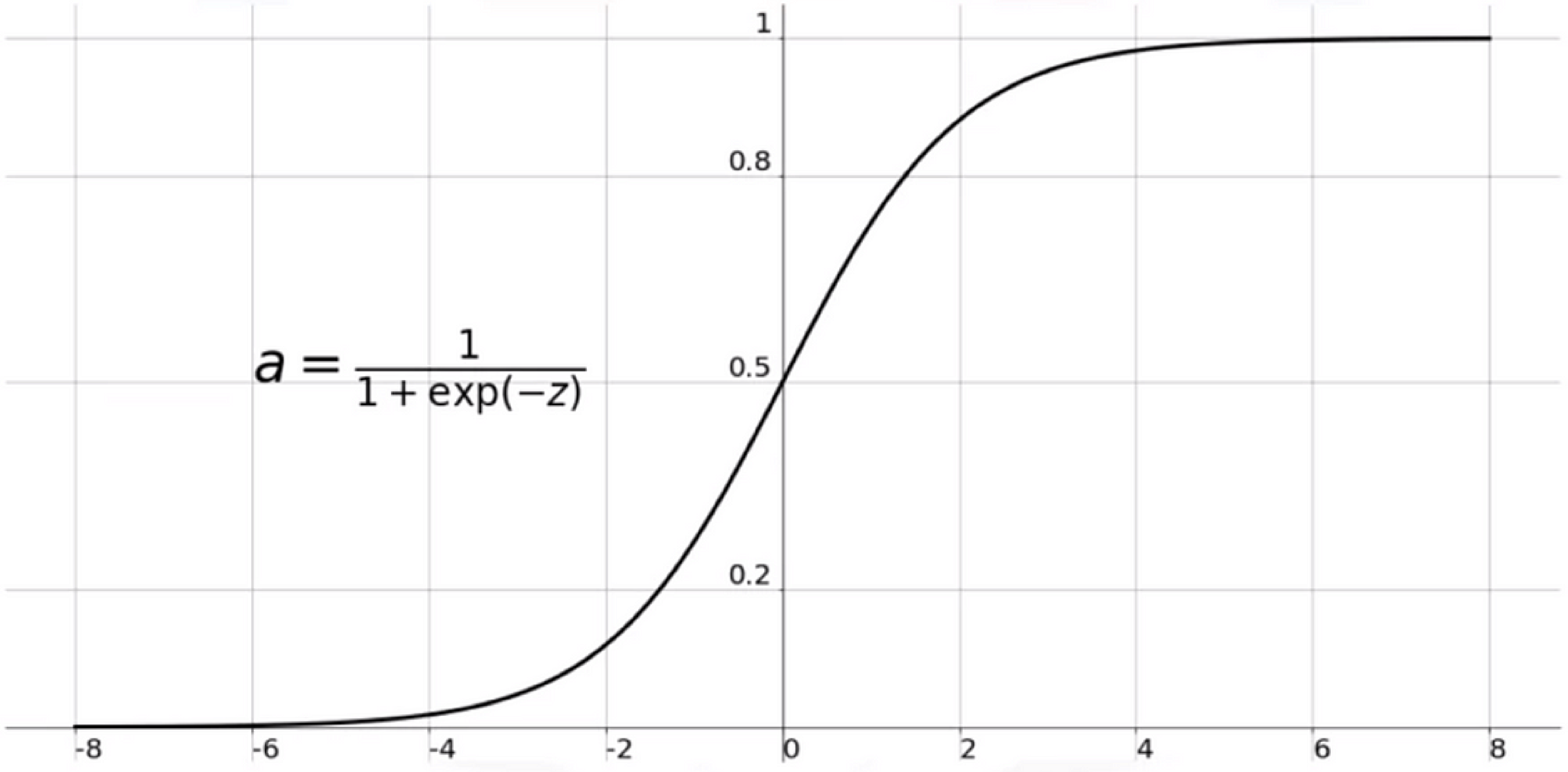
At first, assume a1 and a2 to any value

3. Apply the sigmoid function

$$P(Y = 1| x) = \sigma(z) = h(theta) = 1 / (1 + e^{(-z)})$$

The sigmoid function maps z to a value between 0 and 1, representing the probability of the positive class (passing the exam).

Sigmoid Function



$$a = \frac{1}{1 + \exp(-z)}$$

4. Parameter Estimation

Now we will estimate the parameters a1 and a2 using optimization algorithms.

Cost function for logistic regression:

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}[y^{(i)}\log(h_\theta(x^{(i)})) + (1 - y^{(i)})\log(1 - h_\theta(x^{(i)}))]$$

Updates:

$$a1 = a1 - \alpha * \frac{\partial J}{\partial a1}$$

$$a2 = a2 - \alpha * \frac{\partial J}{\partial a2}$$

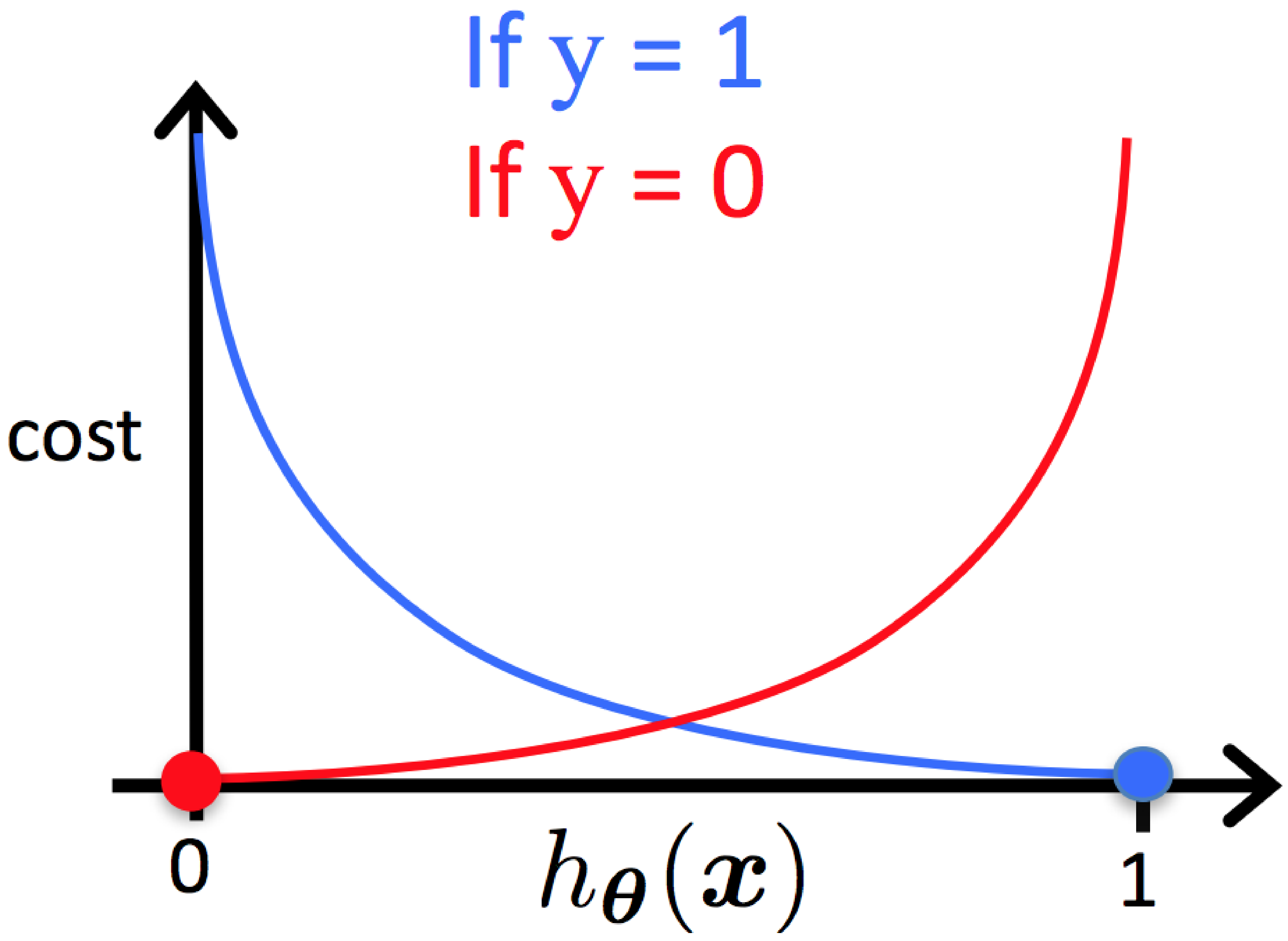
FYI

Cost function for logistic regression is also called Binary cross-entropy or log loss.

FYI

$$\frac{\partial J}{\partial a1} = (h(\theta) - y)x$$

$$\frac{\partial J}{\partial a2} = h(\theta) - y$$


If y = 1
If y = 0
cost
0
$h_\theta(x)$
1

5. Making prediction

Using the estimated parameters, we now make predictions for new examples.

What is the probaility to pass the exam if I study for 4 hours?

z=a1+a2*x

=a1+a2*4

=0.8 (suppose a1=-4 and a2=1.2)

Now, P(Y=1|x=4)=σ(z) = 1 / (1 + e^(-0.8)) ≈ 0.689

6. Binary Classification

Now, we use threshold (0.5) to make binary classification.

Predicted probability > threshold : Passing (1)
Predicted probability < threshold : Failing (0)

Since the obtained probability is 0.689 > 0.5, the student is passing(1).

TechAxis

Congrats! You finished the module!