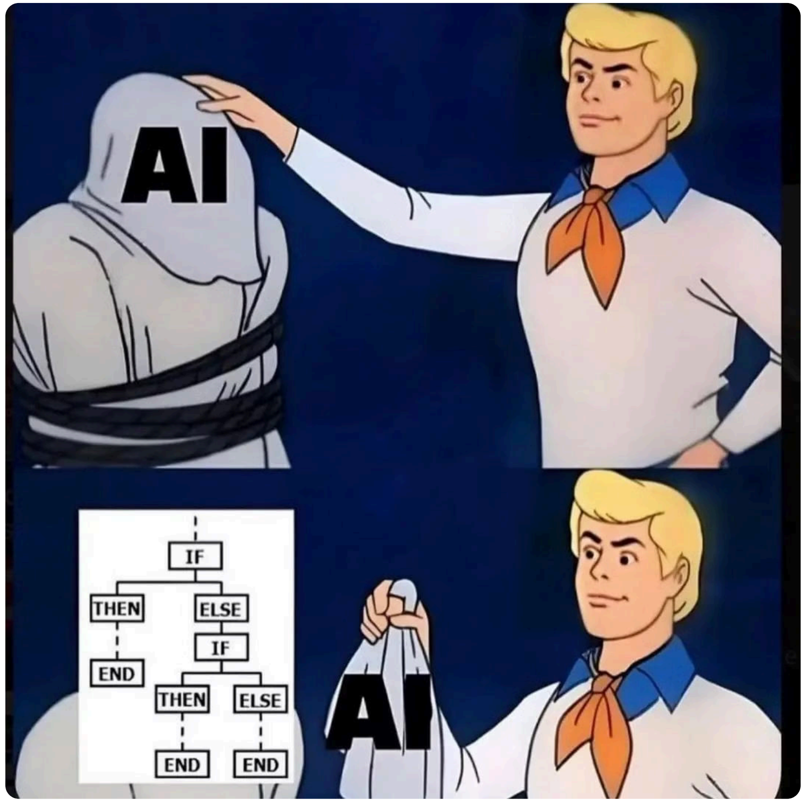
Introduction to Machine Learning

Machine learning is a subfield of artificial intelligence that focuses on the development of algorithms and models that enable computers to learn and make predictions or decisions without being explicitly programmed. It encompasses a wide range of techniques and approaches, including supervised learning, unsupervised learning, reinforcement learning, and more. Understanding the fundamentals of machine learning is crucial in today's technology-driven world, as it forms the backbone of many innovative applications and solutions.

"**machine learning is a field of study that give computers the ability to learn without being explicitly programmed**" – arthur samuel 1959

Traditional Programming vs Machine Learning

In traditional programming, the programmer writes explicit rules and instructions for the computer to follow. With machine learning, the computer learns to recognize patterns and make decisions on its own. While traditional programming is good for tasks with clear rules, machine learning is better suited for tasks with complex patterns or large amounts of data.



Traditional approach :

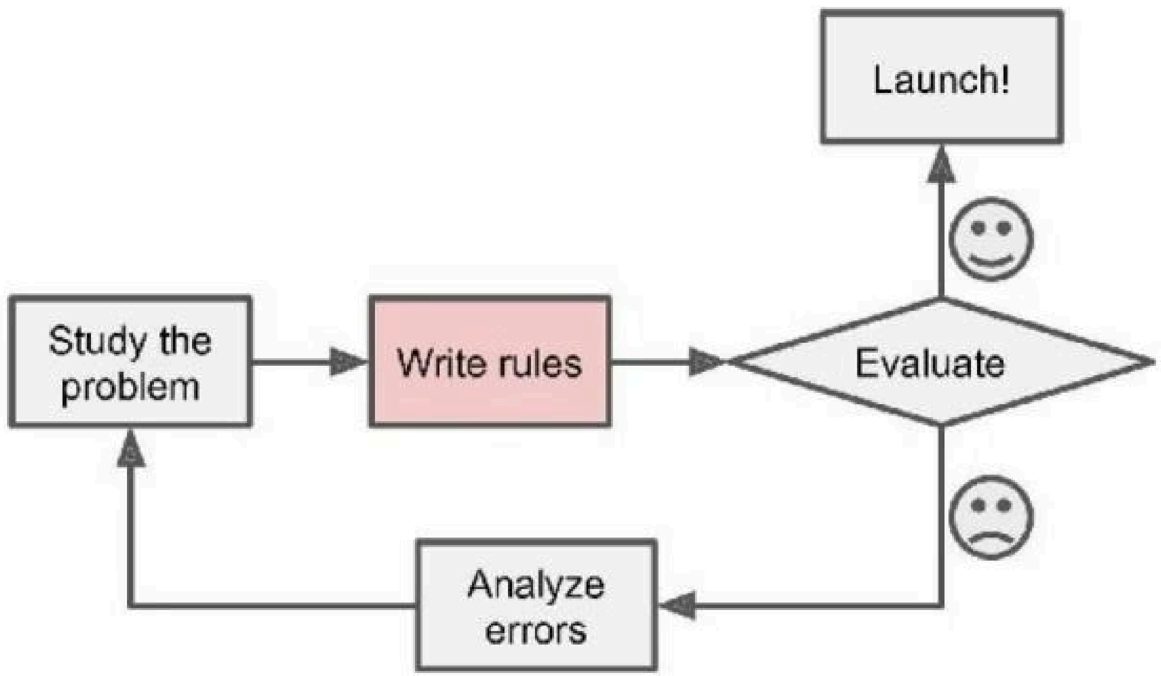


Figure 1-1. The traditional approach

Machine learning approach:

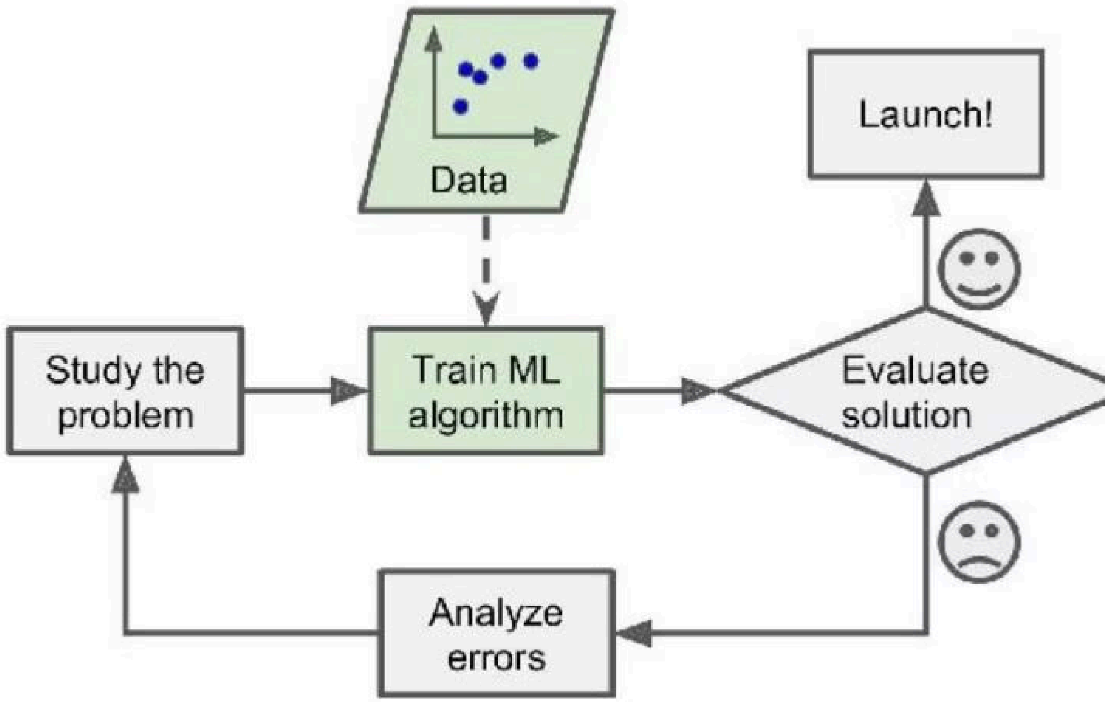
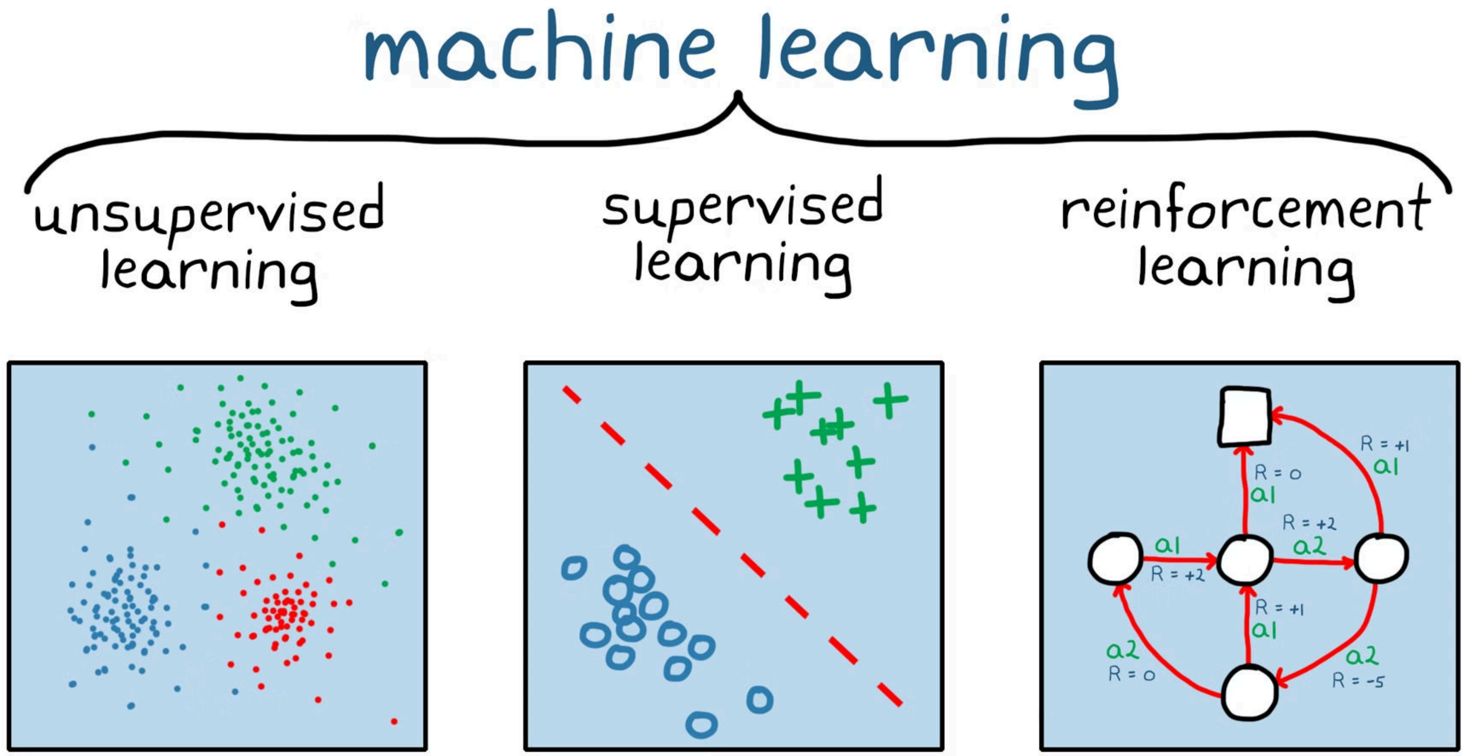
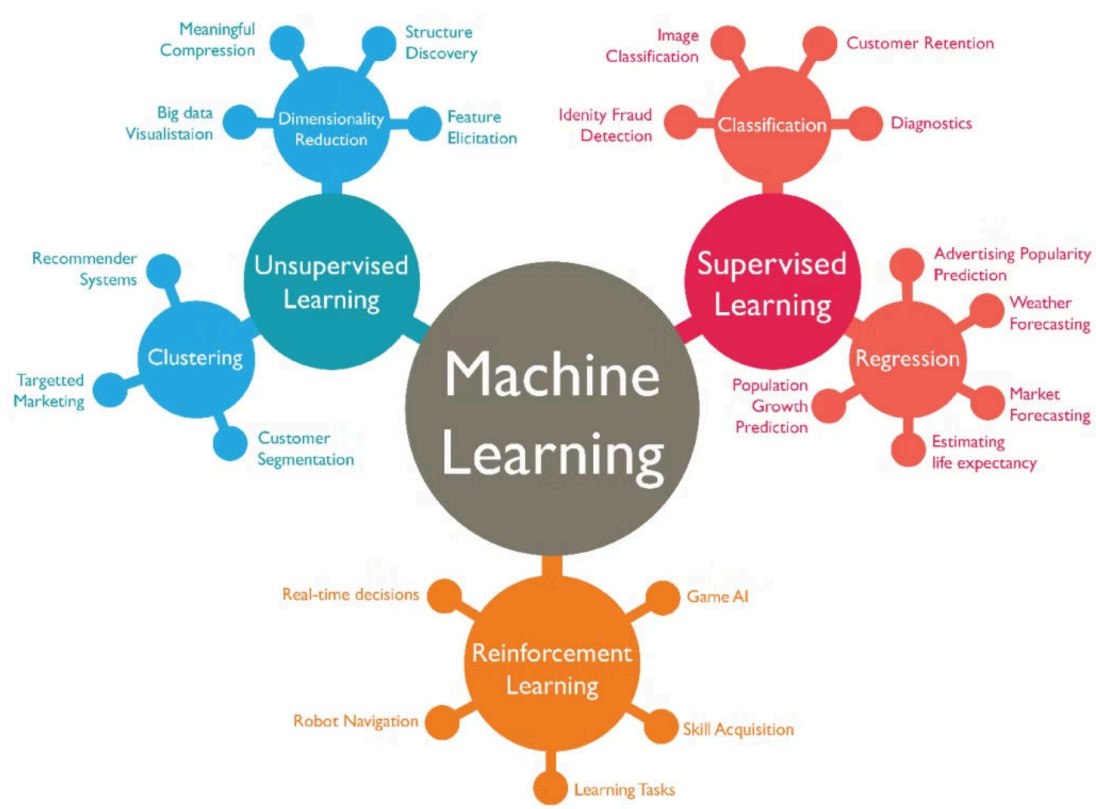


Figure 1-2. Machine Learning approach

Machine Learning





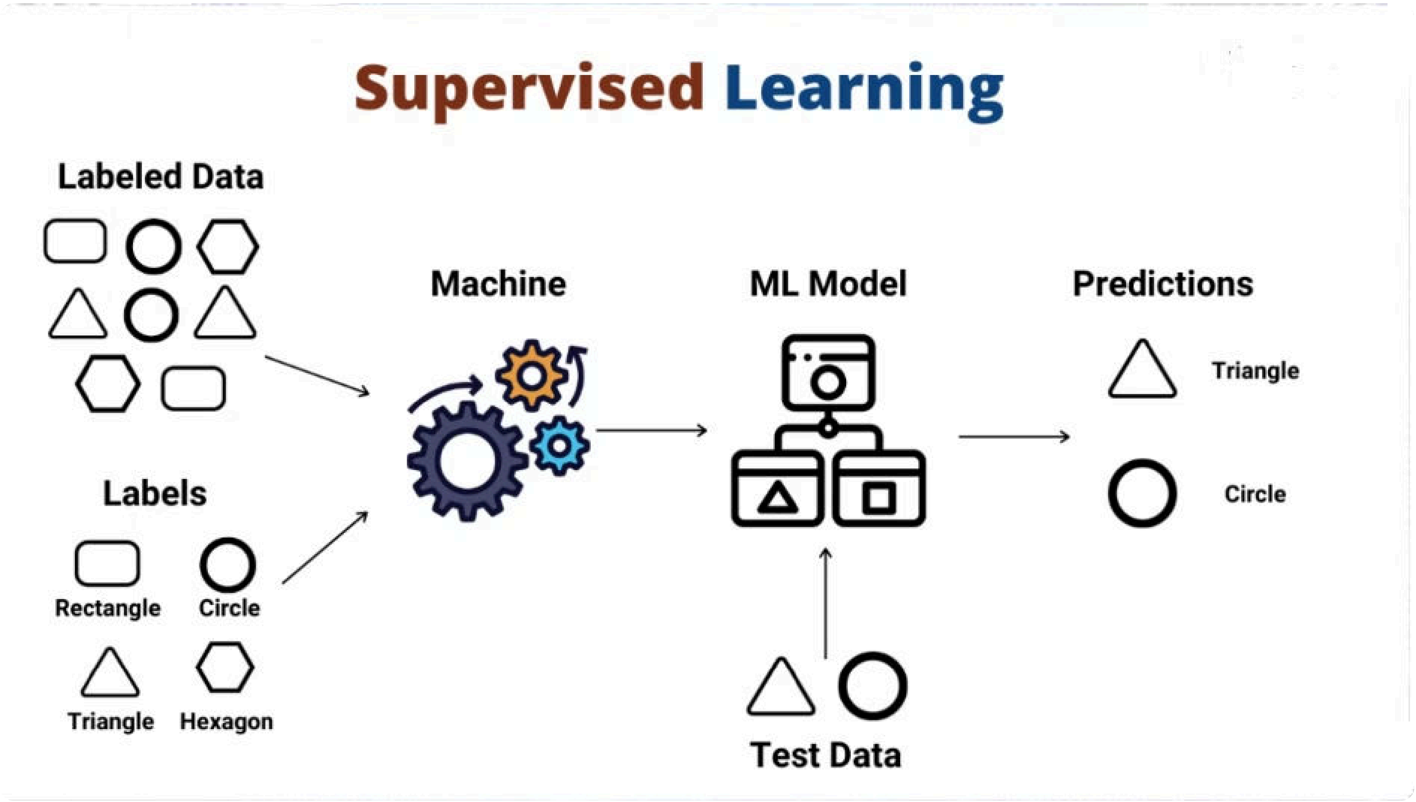
Supervised Learning

Data Labeling

In supervised learning, the algorithm learns from labeled training data, where each input is paired with a corresponding output. This enables the algorithm to make predictions or decisions based on new, unseen data by associating input patterns with target output values.

Regression & Classification

Supervised learning encompasses regression tasks that involve predicting a continuous output and classification tasks that involve categorizing input data into predefined classes or categories.



Supervised Learning
Labeled Data
Labels
Rectangle
Circle
Triangle
Hexagon
Machine
ML Model
Test Data
Predictions
Triangle
Circle

Unsupervised Learning

1 Clustering & Association

Unsupervised learning focuses on identifying patterns and structure within input data, including techniques such as clustering to group similar data points and association to discover relationships among variables.

2 Anomaly Detection

Another key aspect of unsupervised learning is anomaly detection, which aims to identify unusual or abnormal patterns in the data that may require further investigation or action.

Reinforcement Learning

1

Agent-Environment Interaction

Reinforcement learning involves an agent interacting with an environment, learning to achieve a goal through the consequence of its actions, and receiving feedback in the form of rewards or penalties.

2

Exploration & Exploitation

One of the key challenges in reinforcement learning is finding the right balance between exploration (trying new actions to discover potential rewards) and exploitation (leveraging known actions for immediate rewards).

Regression: Logistic, Linear

Linear Regression

Linear regression models the relationship between a dependent variable and one or more independent variables by fitting a linear equation to the observed data.

Logistic Regression

Logistic regression is used for binary classification tasks, where the output is a probability estimate for a particular class.

Cost Function

Minimization

Minimization

The cost function aims to be minimized during the training process to find optimal model parameters that best fit the data.

Optimization Techniques

Optimization Techniques

Various techniques such as gradient descent are used to iteratively minimize the cost function and update model parameters.

cost function

$$J(w,b) = \frac{1}{2m}\sum_{i=1}^{m}(f_{w,b}(x^{(i)}) - y^{(i)})^2$$

Gradient Descent

1 Initialization

Gradient descent starts with the initialization of model parameters or weights with certain values.

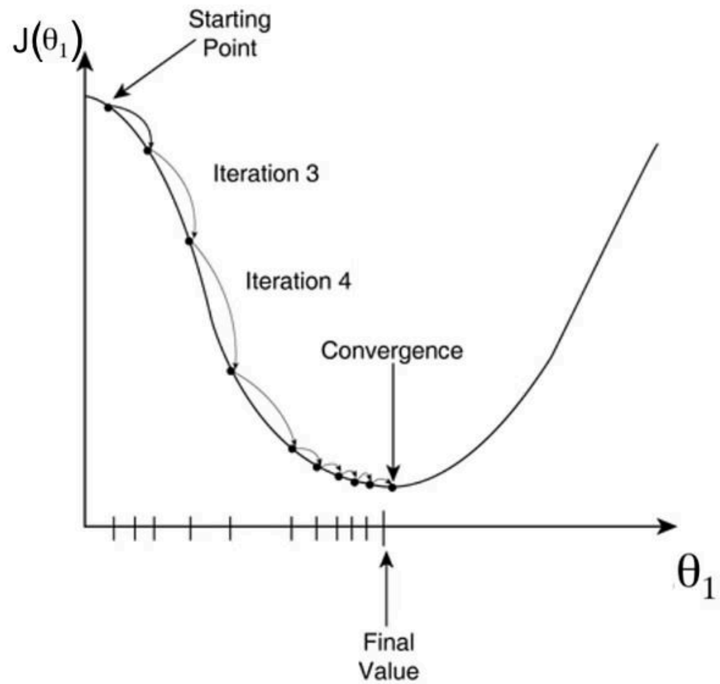
2 Update

During each iteration, the model parameters are updated in the direction that minimizes the cost function, guided by the gradient of the function with respect to the parameters.

3 Convergence

Gradient descent continues the update process until it converges to the minimum value or a predefined stopping criterion is met.

Example : linear regression



Cost Function – "One Half Mean Squared Error":

$$J(\theta_0, \theta_1) = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

Objective:

$$\min_{\theta_0, \theta_1} J(\theta_0, \theta_1)$$

Derivatives:

$$\frac{\partial}{\partial\theta_0} J(\theta_0, \theta_1) = \frac{1}{m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right)$$

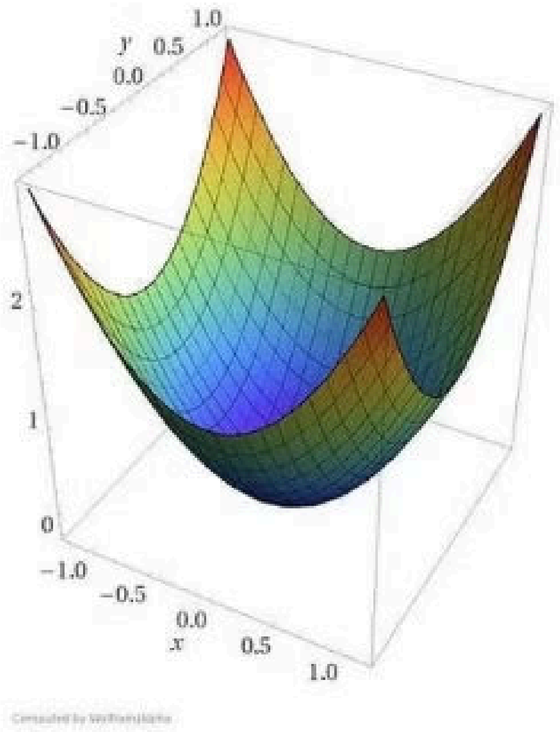
$$\frac{\partial}{\partial\theta_1} J(\theta_0, \theta_1) = \frac{1}{m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right) \cdot x^{(i)}$$

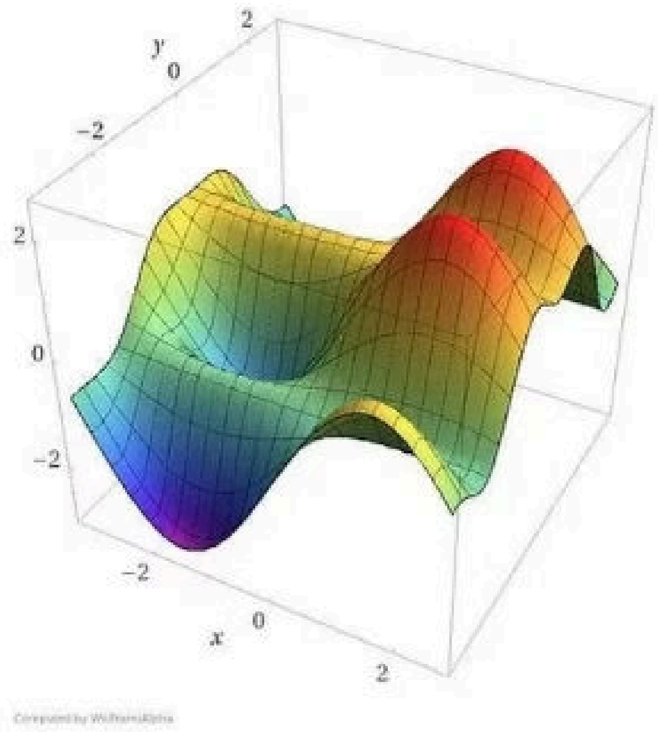
Repeat until convergence {

$$\theta_1 \leftarrow \theta_1 - \alpha \, \frac{1}{m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right)$$

$$\theta_2 \leftarrow \theta_2 - \alpha \, \frac{1}{m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right) \cdot x^{(i)}$$

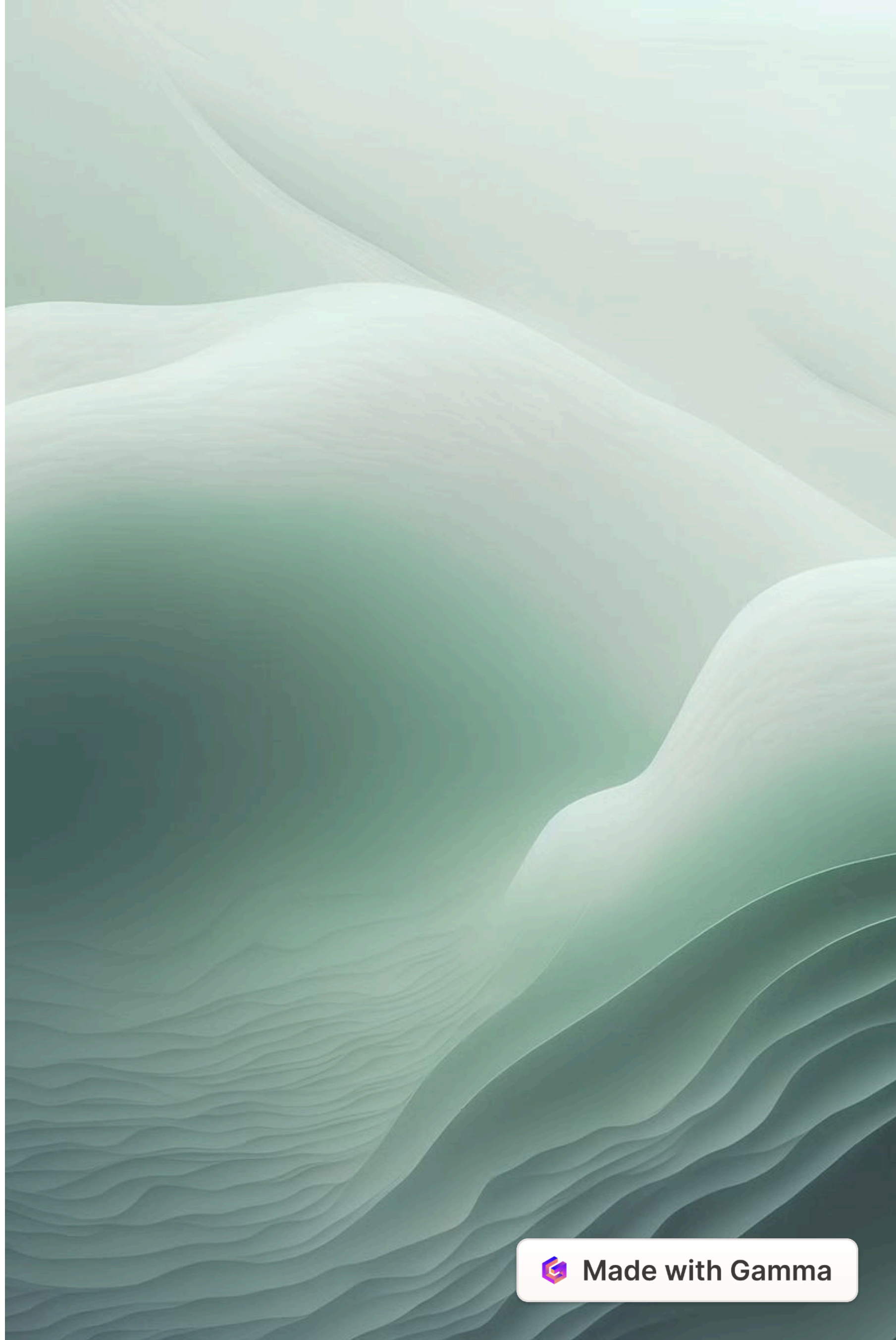
}


y
1.0
0.5
0.0
−0.5
−1.0
2
1
0
x
−1.0
−0.5
0.0
0.5
1.0


y
2
0
−2
2
0
−2
x
−2
0
2

Gradient Descent Example: Linear Regression

In linear regression, gradient descent is used to minimize the cost function and find the optimal values for the regression coefficients. By iteratively adjusting the coefficients based on the gradient of the cost function, the algorithm converges to the best fit line that minimizes the error between the predicted and actual values.



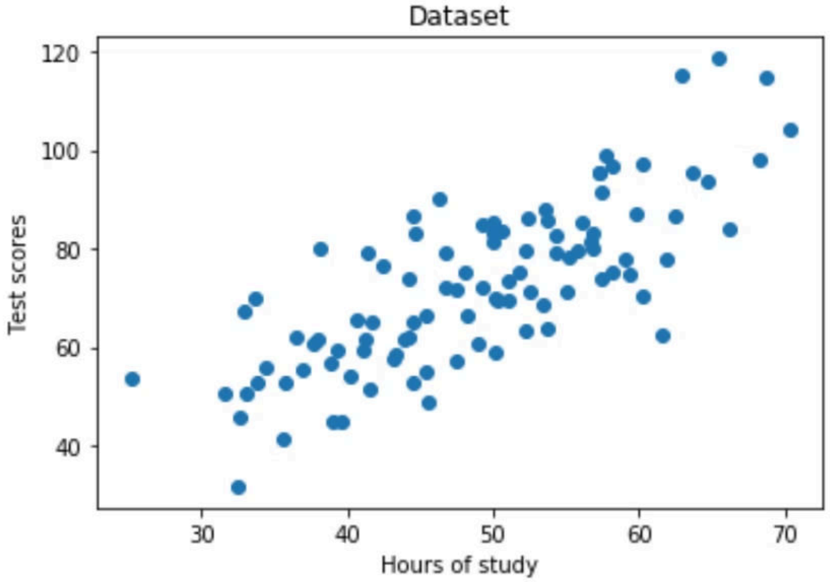
Example : linear regression

Suppose we want to predict the students score bases on study hours.

- Mathematical model

$$y' = wx + b$$

y : the output (score predicted)
x : the input (study hours).
w : weight
b : bias



Example : linear regression

Linear regression model

$$f_{w,b}(x) = wx + b$$

Cost function

$$J(w,b) = \frac{1}{2m}\sum_{i=1}^{m}(f_{w,b}(x^{(i)}) - y^{(i)})^2$$

$$w = w - \alpha\,\frac{1}{m}\sum_{i=1}^{m}(f_{w,b}(x^{(i)}) - y^{(i)})x^{(i)}$$

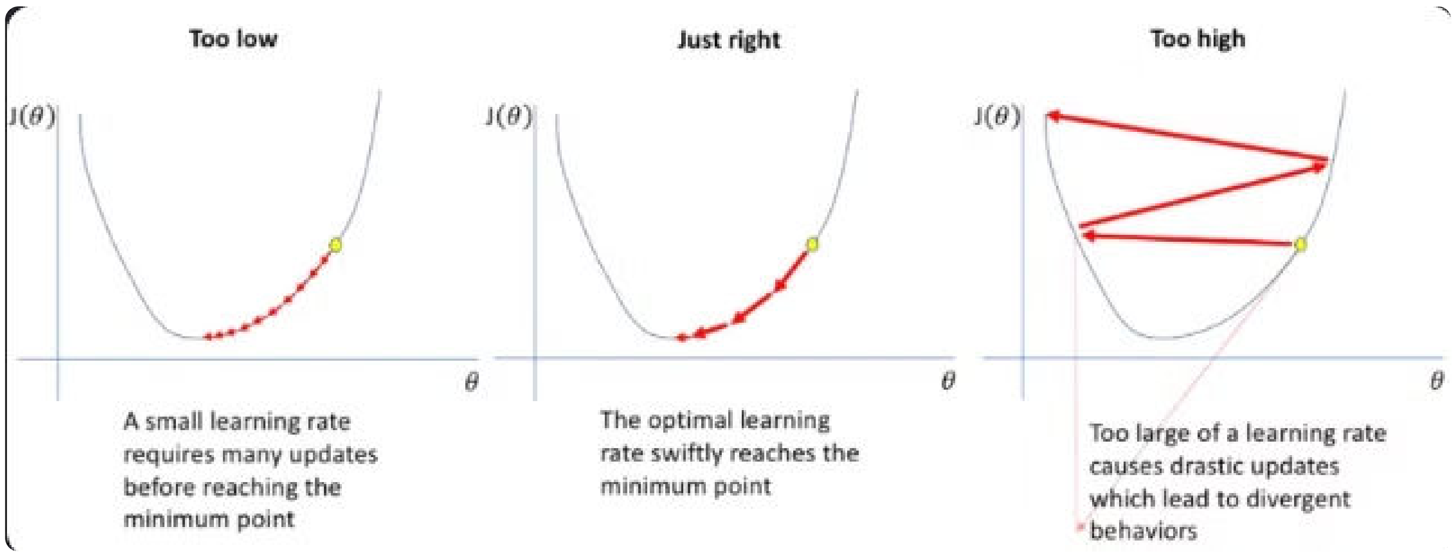
$$b = b - \alpha\,\frac{1}{m}\sum_{i=1}^{m}(f_{w,b}(x^{(i)}) - y^{(i)})$$

Example : linear regression

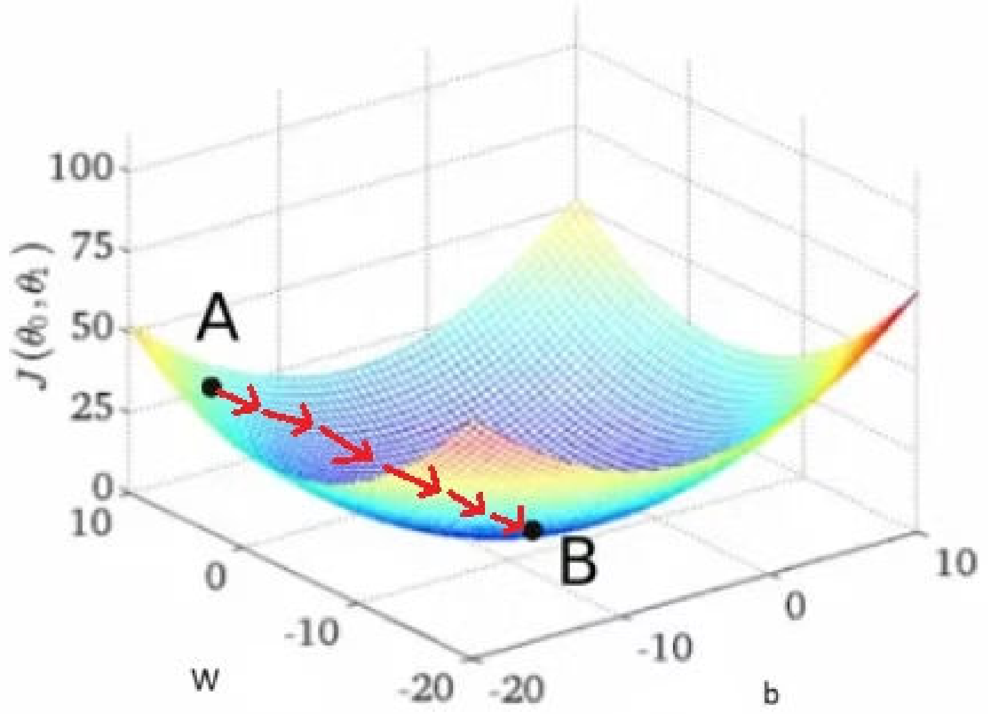
Learning rate α : is a hyperparameter that controls how much to change the model in response to the estimated error each time the model weights are updated.

$$w = w - \alpha \frac{1}{m} \sum_{i=1}^{m} (f_{w,b}(x^{(i)}) - y^{(i)}) x^{(i)}$$

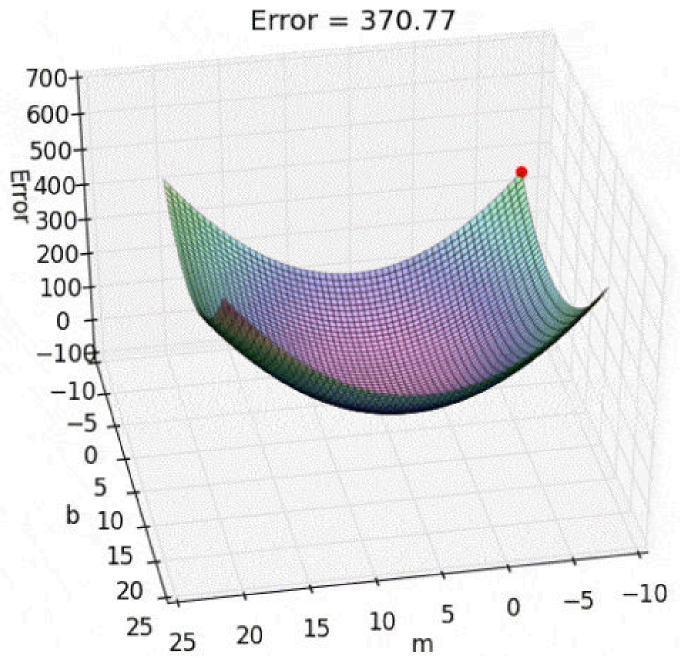
$$b = b - \alpha \frac{1}{m} \sum_{i=1}^{m} (f_{w,b}(x^{(i)}) - y^{(i)})$$

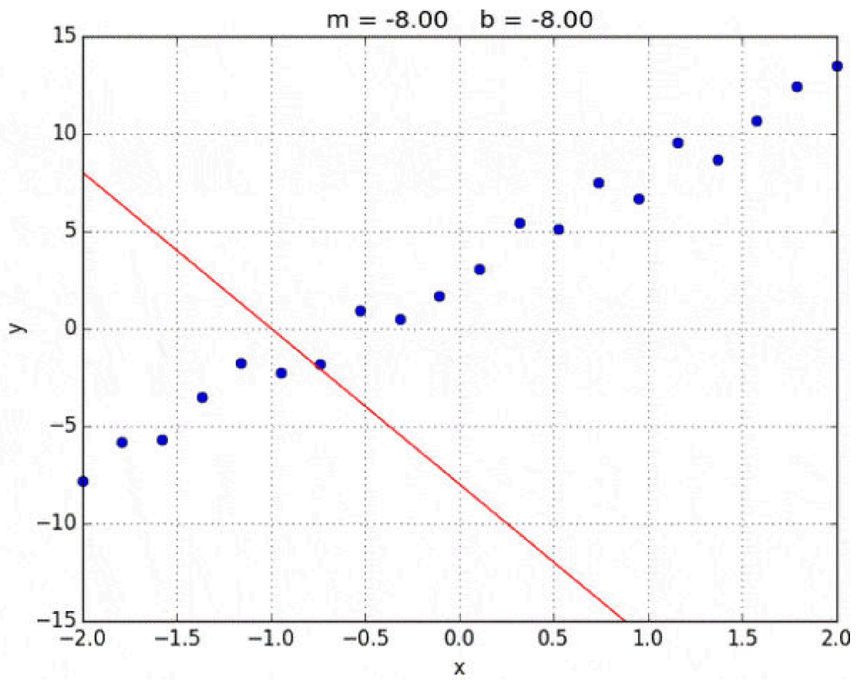


Example : linear regression

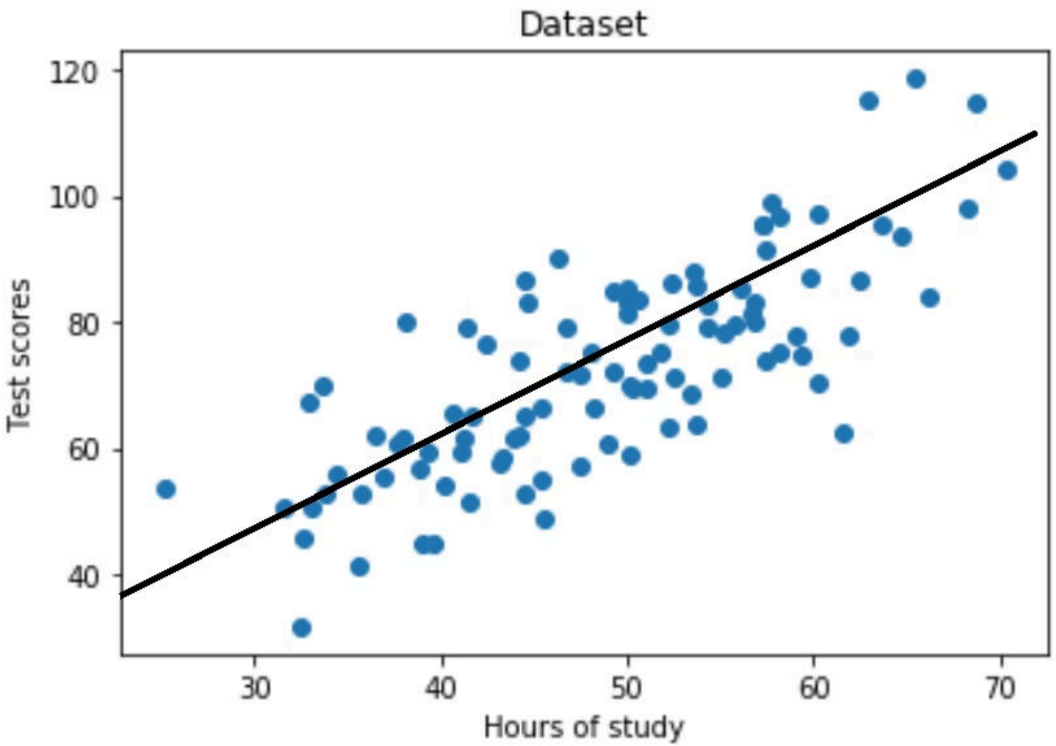


Example : linear regression

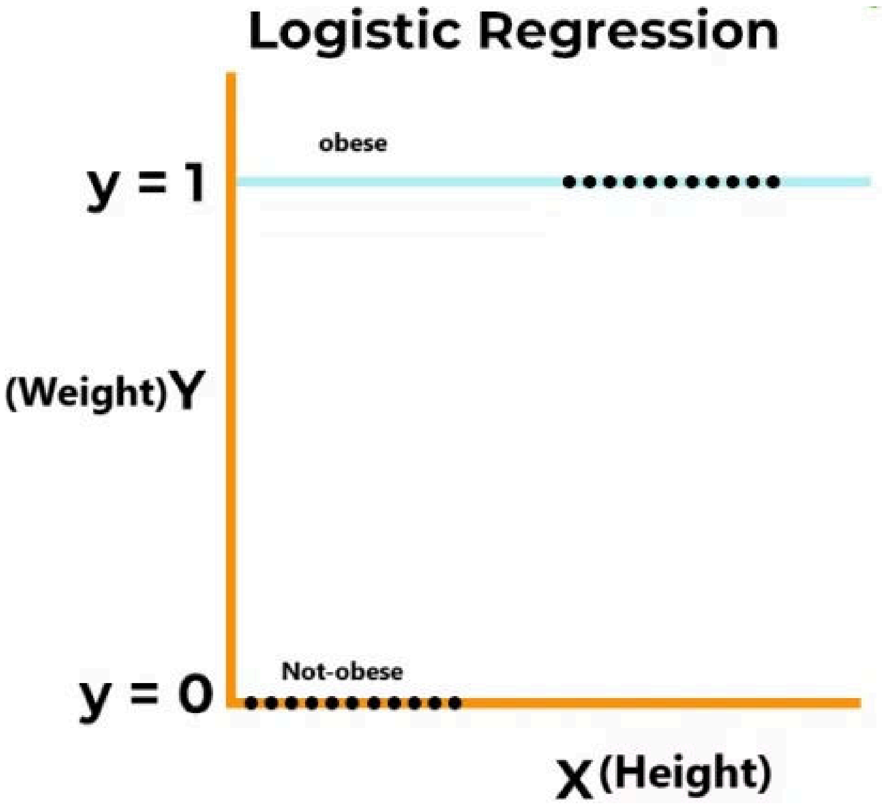




Example : linear regression

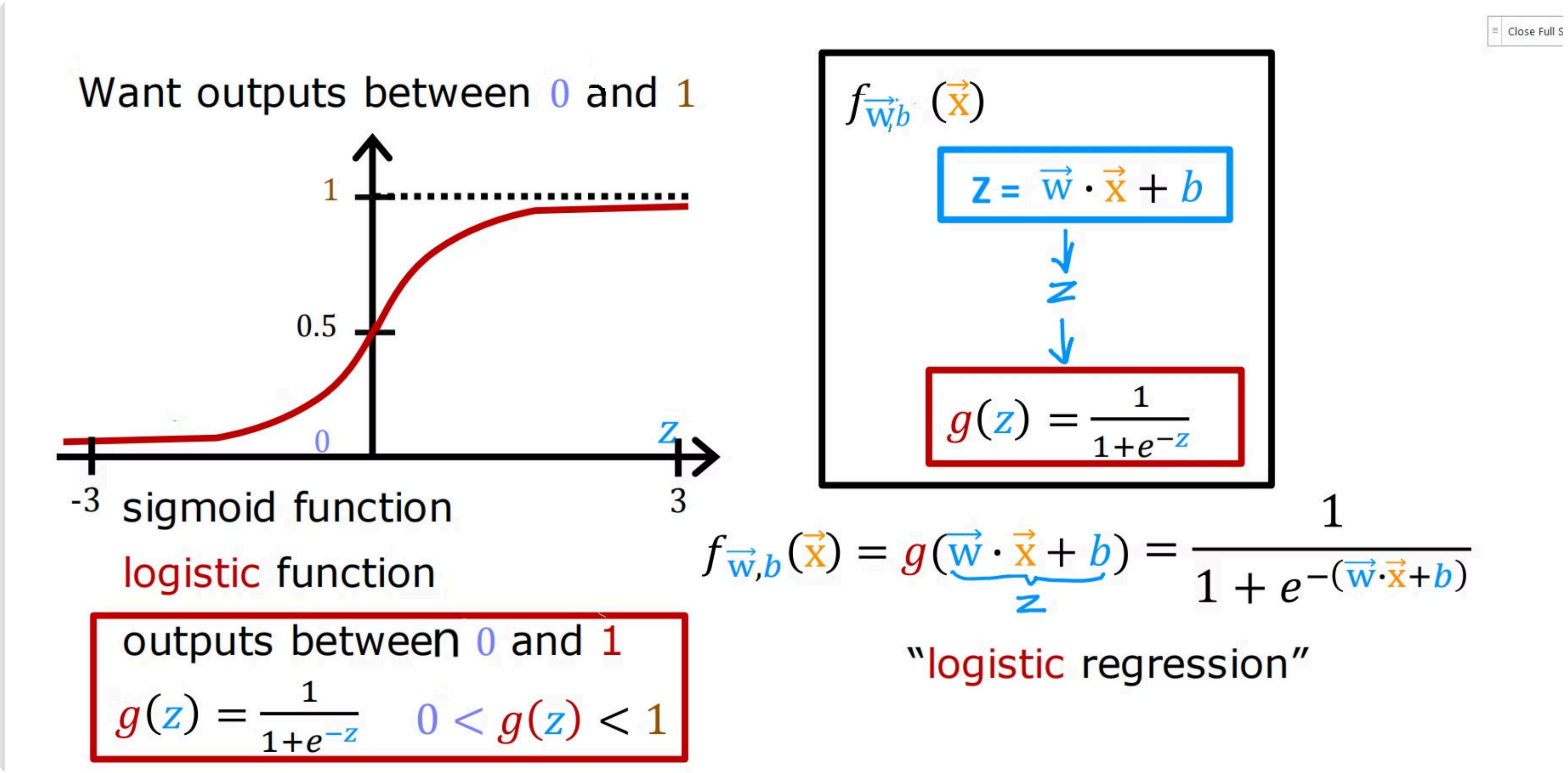


Example : logistic regression



Example : logistic regression

Sigmoid Function (logistic function) :



Example : logistic regression

$$L\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right), y^{(i)}\right) = \begin{cases} \boxed{-\log\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right)\right) \quad \text{if } y = 1} \\ \boxed{-\log\left(1 - f_{\vec{w},b}\left(\vec{x}^{(i)}\right)\right) \quad \text{if } y = 0} \end{cases}$$

Simplified cost function

$$L\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right), y^{(i)}\right) = -y^{(i)}\log\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right)\right) - \left(1 - y^{(i)}\right)\log\left(1 - f_{\vec{w},b}\left(\vec{x}^{(i)}\right)\right)$$

$$J(\vec{w}, b) = \frac{1}{m}\sum_{i=1}^{m}\left[L\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right), y^{(i)}\right)\right]$$

$$= -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\log\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right)\right) + \left(1 - y^{(i)}\right)\log\left(1 - f_{\vec{w},b}\left(\vec{x}^{(i)}\right)\right)\right]$$

Example : logistic regression

Gradient descent for logistic regression

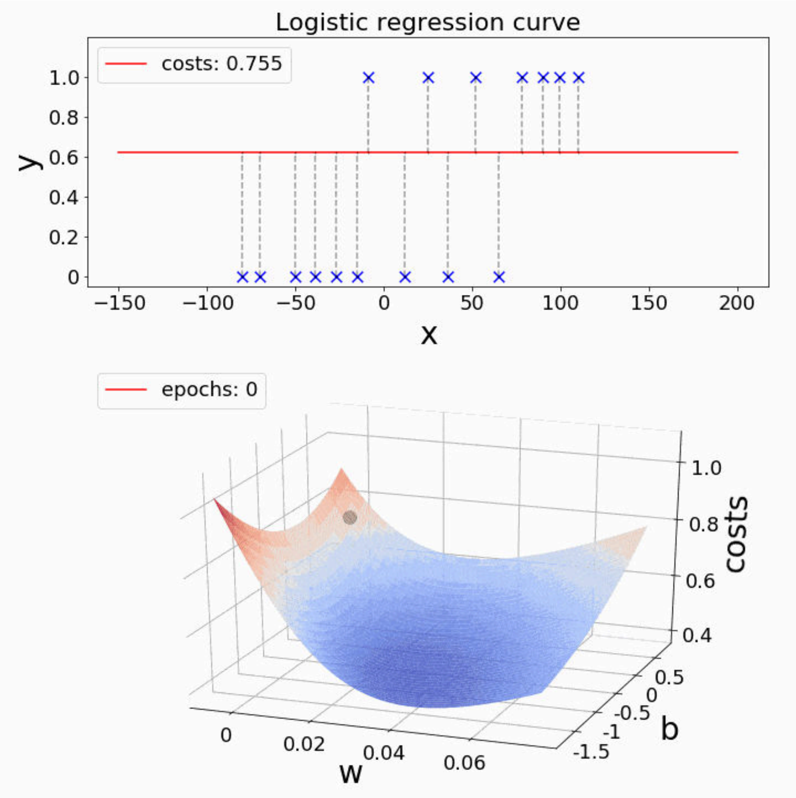
repeat {

$$w_j = w_j - \alpha \left[\frac{1}{m} \sum_{i=1}^{m} \left(f_{\vec{\mathrm{w}},b}\left(\vec{\mathrm{x}}^{(i)}\right) - y^{(i)} \right) x_j^{(i)} \right]$$

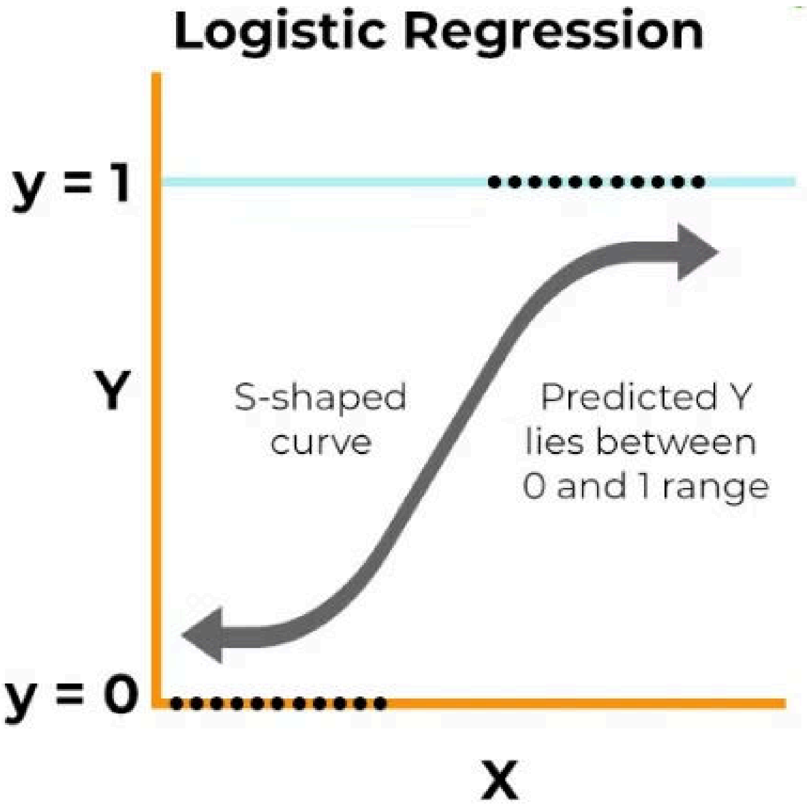
$$b = b - \alpha \left[\frac{1}{m} \sum_{i=1}^{m} \left(f_{\vec{\mathrm{w}},b}\left(\vec{\mathrm{x}}^{(i)}\right) - y^{(i)} \right) \right]$$

} simultaneous updates

Example : logistic regression

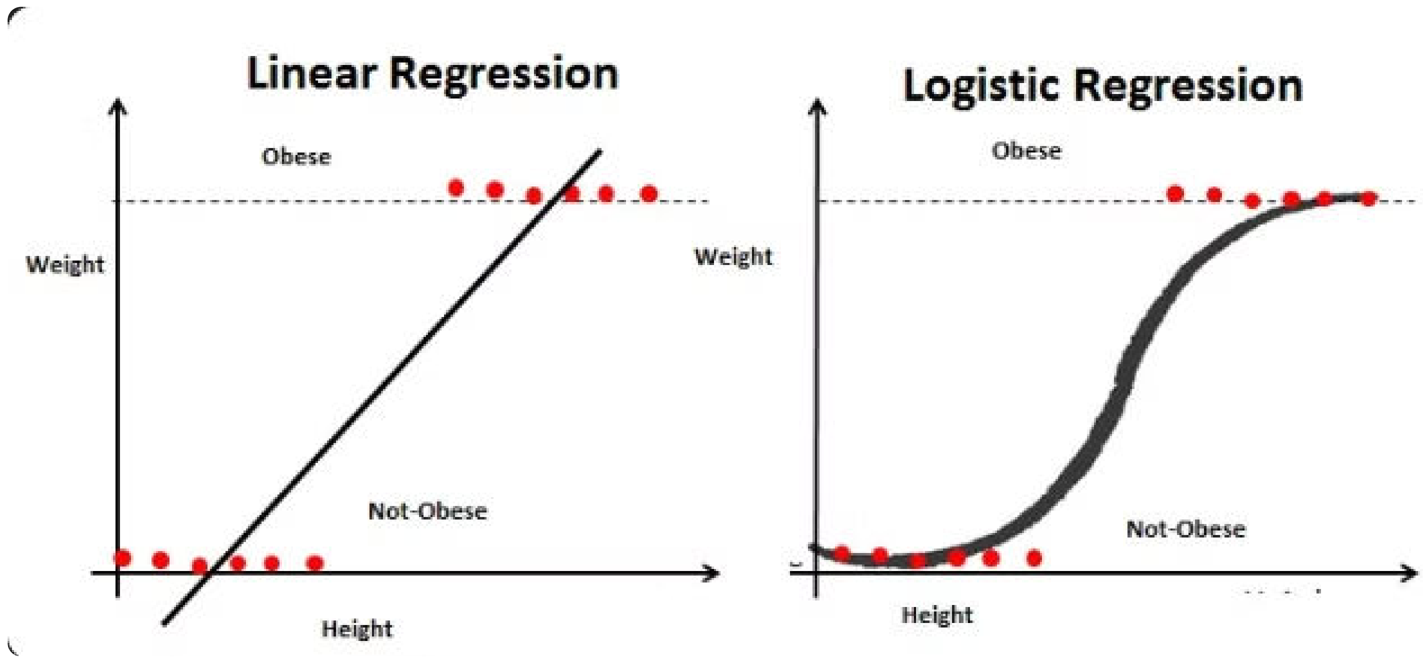


Example : logistic regression



Made with Gamma

logistic regression vs linear regression



Linear regression $f_{\vec{w},b}(\vec{x}) = \vec{w} \cdot \vec{x} + b$

Logistic regression $f_{\vec{w},b}(\vec{x}) = \dfrac{1}{1 + e^{(-\vec{w}\cdot\vec{x}+b)}}$