Machine Learning Bootcamp

By DRCFS and Mercantile Cloud

What is Data Research Council for Students?

- Non-profit organization focused on data literacy and student support.
- Started with two students, now Nepal's largest data community.
- Believes data literacy is essential for success in the modern world.
- Offers workshops, seminars, training, and hackathon for students.
- Provides collaborative and supportive environment for learning.
- Aims to build a brighter future for Nepal and beyond through

What is Machine Learning

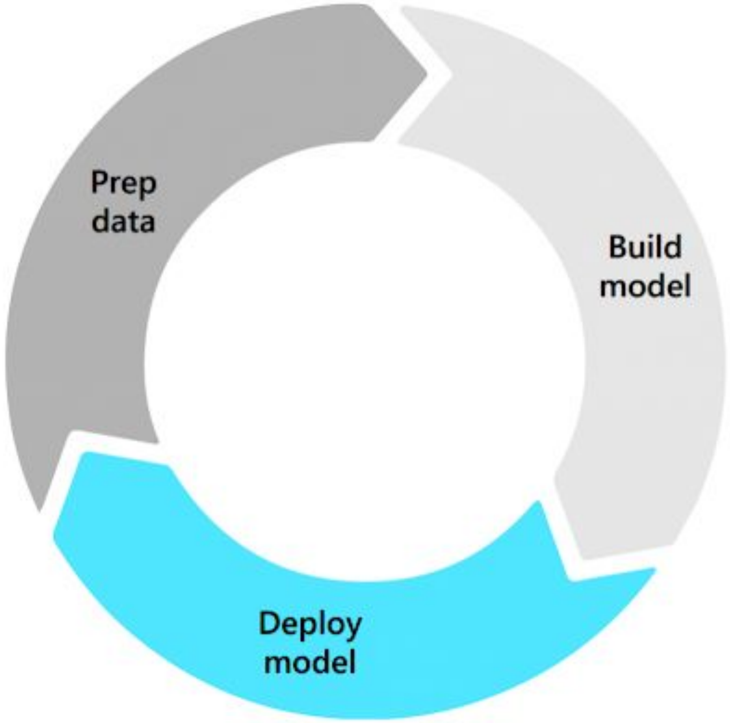
What is Machine Learning

मेशिन लर्निङ = ठुलिपोखरीको माछा

मेशिन लर्निङ = ठुलिपोखरीको माछा

मेशिन लर्निङ = ठुलिपोखरीको माछा

Life cycle of any machine learning project



Types of machine learning algos

1. Supervised machine learning
2. Unsupervised machine learning
3. Reinforcement learning

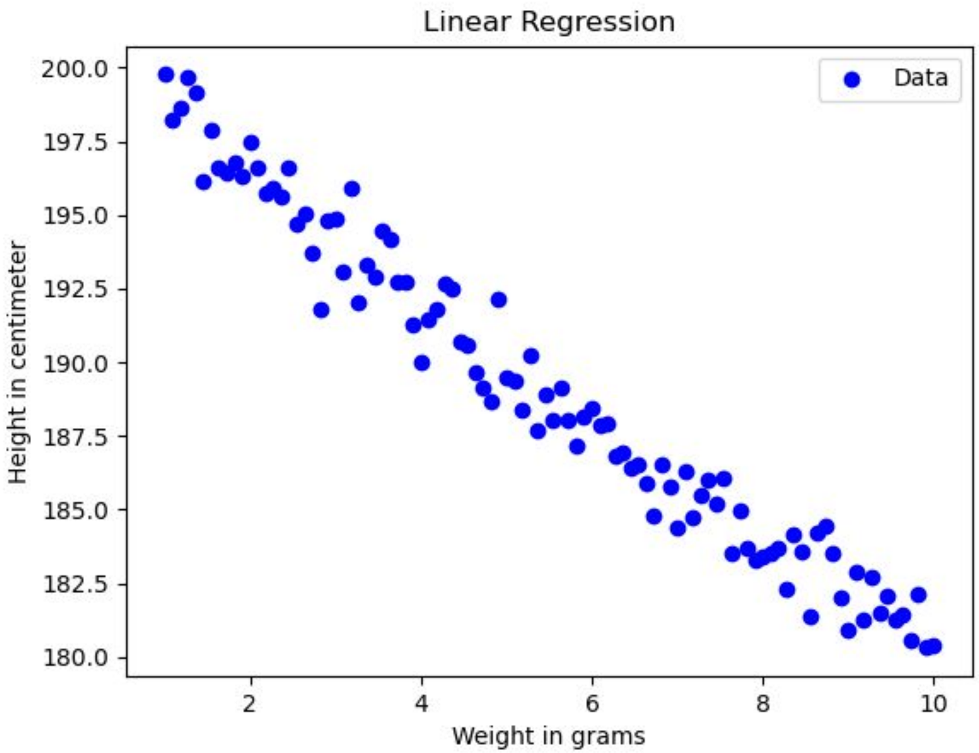
Supervised Machine learning

- The fishes learn the hand gestures and associate them with food availability
- Model learns from the labeled data (hand gesture -> food, no hand gesture -> no food) and recognizes pattern
- Any problem where the label of the data is known and we train our model according to the labelled data is supervised machine learning

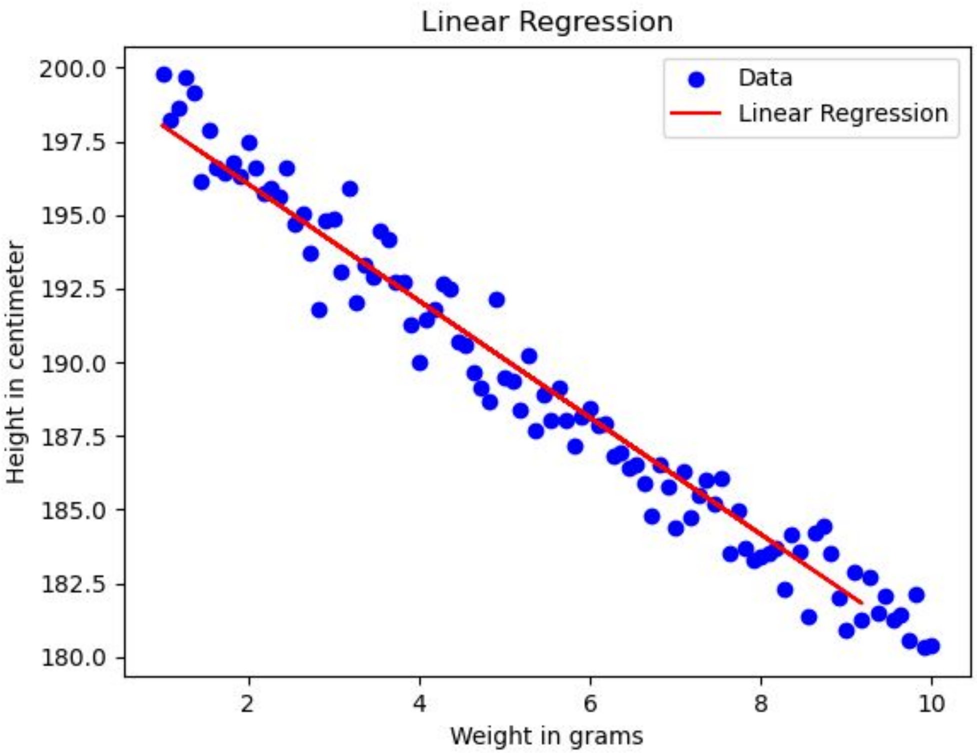
Linear Regression

Imagine you have a collection of balloons with different weights and corresponding heights. Your goal is to understand the relationship between the weight of a balloon and its height, and use that understanding to predict the height of new balloons based on their weights.

Linear Regression



Linear Regression



Linear Regression mathematically

In the simplest form

$Y = mx + b$ is a linear regression for a single variable input and output

For multiple variables:

$Y = w.x+b$,

Where x and Y are vectors that encapsulate all the variables.

How to find the best ‘w’ and ‘b’ for any linear regression?

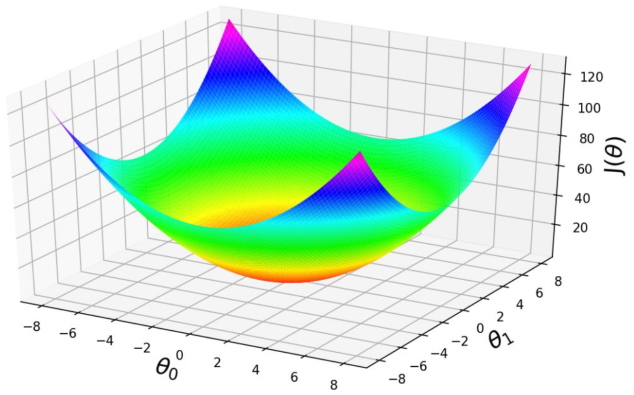
Step 1	Step 2	Step 3
Start with random w and b and find how off our model is using something called a loss function or error function	Use the loss function or error function to update our w and b such that the we move towards the minimum of the loss function	Use the updated w and b to calculate our loss again and repeat step 2 until convergence

What is loss function?

- It measures how well the model is performing in terms of its predictions.
- It helps the model adjust its parameters to minimize the discrepancy between predicted and actual values.
- Mean Squared Error (MSE): Commonly used for regression problems, it calculates the average squared difference between predicted and actual values.

$$\mathrm{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \tilde{y}_i)^2$$

How to reduce loss?



Gradient Descent

- It calculates the gradient of the loss function with respect to the parameters and updates them in the direction of steepest descent.
- It isolates w and b to update them individually while considering the other value as constant
- Consider theta0 and theta1 as the w and b in the given case

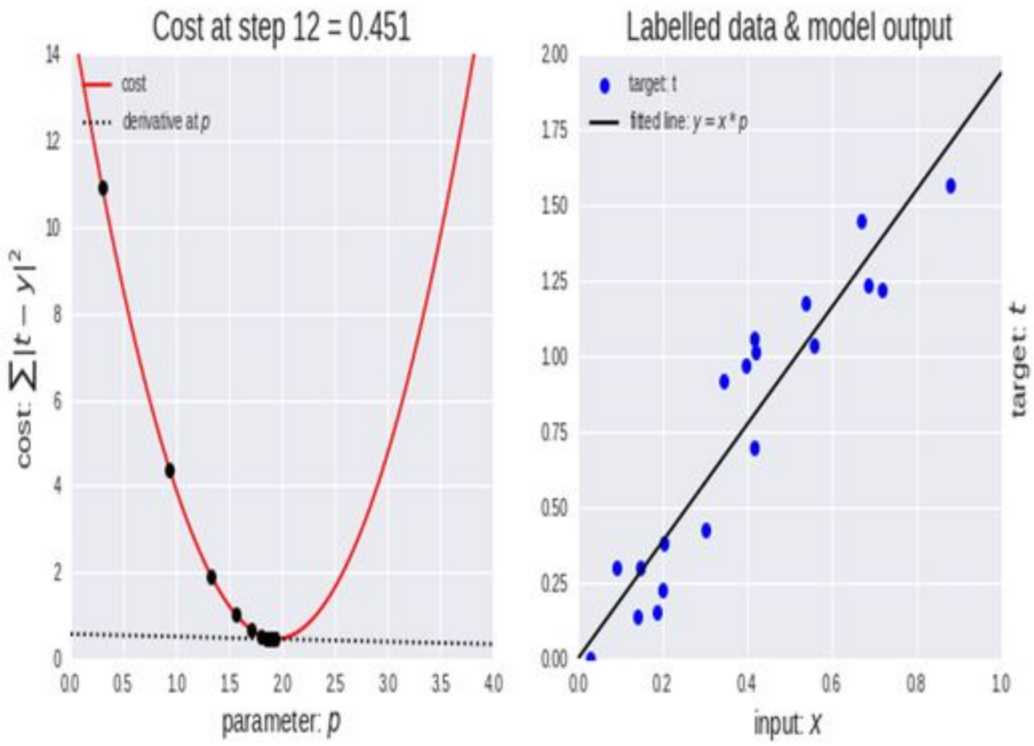
Gradient descent mathematically

- Take slope of the loss function individually twice, once with respect to m and once with b
- Calculate the gradient and move towards a lower point of the steep

$$\frac{\partial}{\partial \mathrm{m}} = \frac{2}{N} \sum_{i=1}^{N} -x_i(y_i - (mx_i + b))$$

$$\frac{\partial}{\partial \mathrm{b}} = \frac{2}{N} \sum_{i=1}^{N} -(y_i - (mx_i + b))$$

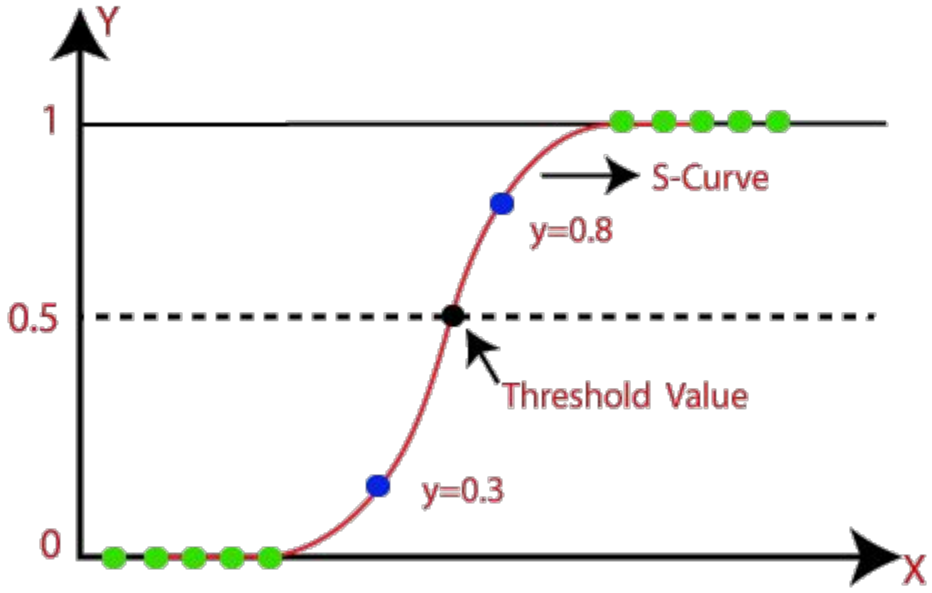
Gradient descent visually



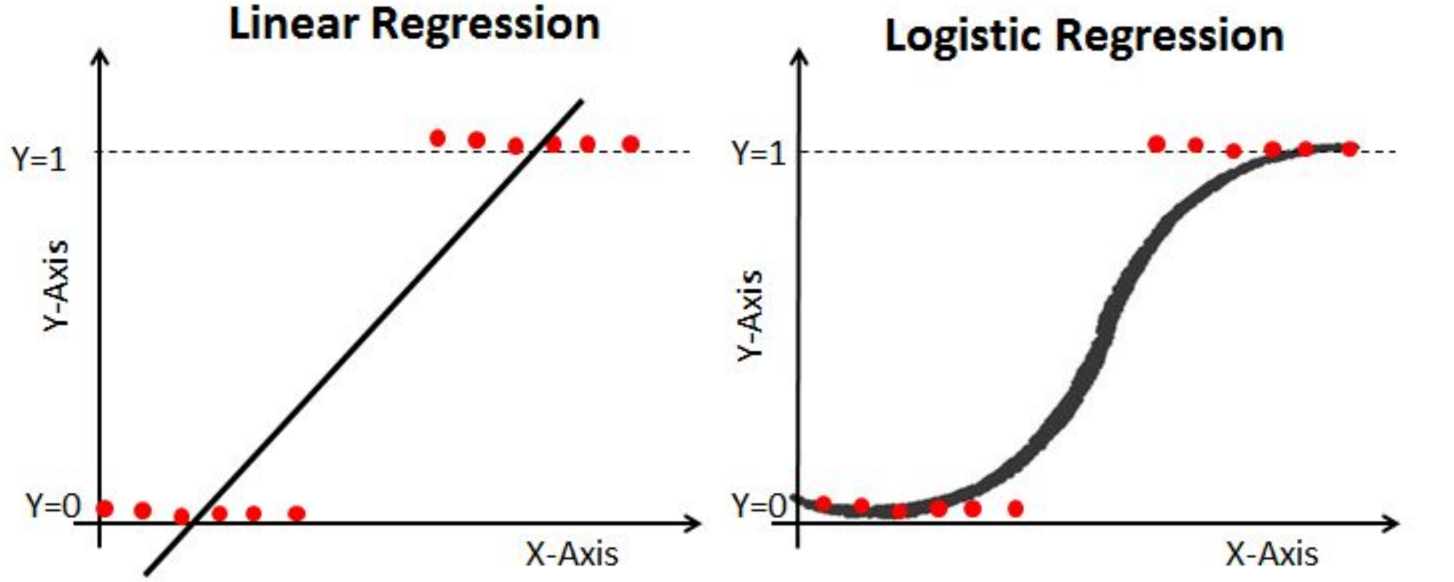
But what if?

- What if the data has y label values as boolean values instead of linearly increasing values.
- Suppose in dataset that classifies breast tumor as malignant or benign the data would be either 0 or 1 (B or M)

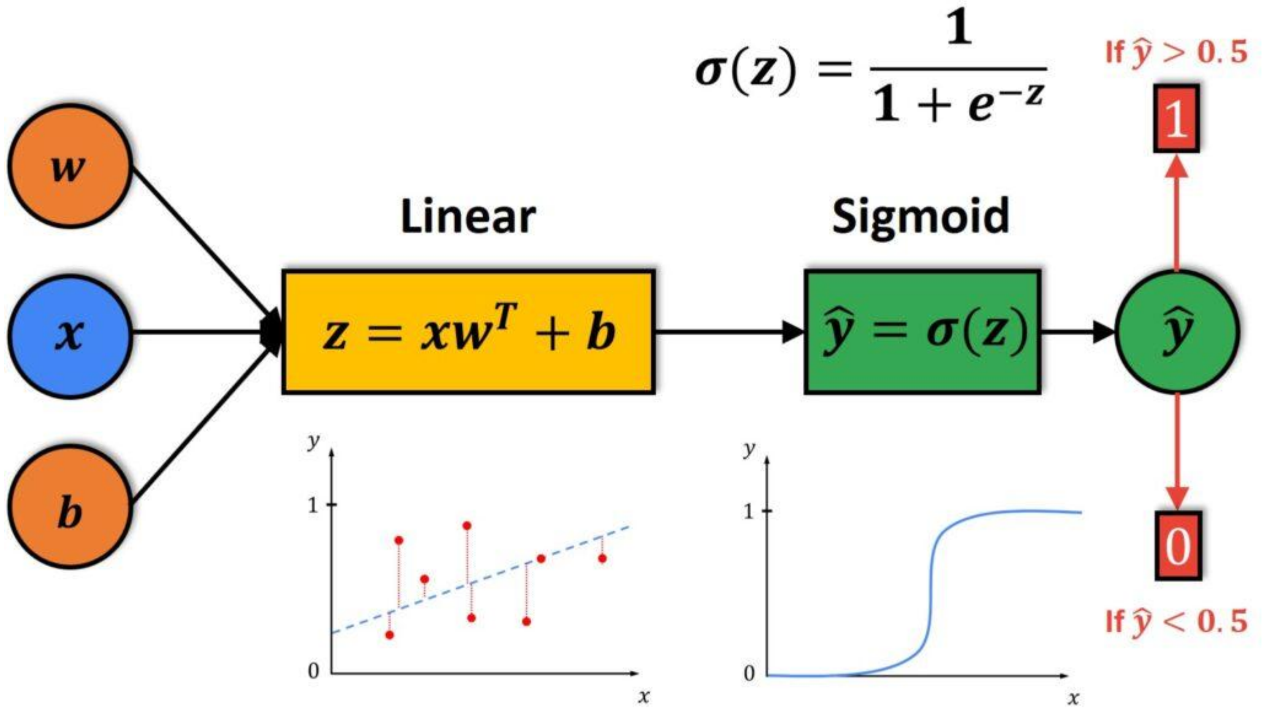
Logistic Regression



Linear Regression vs Logistic Regression



Logistic Regression mathematically



Loss function in logistic regression

Y is the real label and y(hat) is the predicted probability.

- It measures the dissimilarity between predicted probabilities and true binary labels.
- By minimizing the loss, the model learns to predict probabilities that align closely with the true binary labels, improving its ability to classify new instances accurately.

$$L_{BCE} = -\frac{1}{n}\sum_{i=1}^{n}(Y_i \cdot \log \hat{Y}_i + (1 - Y_i) \cdot \log(1 - \hat{Y}_i))$$

Gradient Descent in BCE

$$\mathbf{h} = \mathbf{w}^T\mathbf{X}$$

$$\text{Logistic regression: } \mathbf{z} = \sigma(\mathbf{h}) = \frac{1}{1+e^{-\mathbf{h}}}$$

$$\text{Cross-entropy loss: } J(\mathbf{w}) = -(\mathbf{y}log(\mathbf{z}) + (1-\mathbf{y})log(1-\mathbf{z}))$$

$$\text{Use chain rule: } \frac{\partial J(\mathbf{w})}{\partial \mathbf{w}} = \frac{\partial J(\mathbf{w})}{\partial \mathbf{z}}\frac{\partial \mathbf{z}}{\partial \mathbf{h}}\frac{\partial \mathbf{h}}{\partial \mathbf{w}}$$

$$\frac{\partial J(\mathbf{w})}{\partial \mathbf{z}} = -(\frac{\mathbf{y}}{\mathbf{z}} - \frac{1-\mathbf{y}}{1-\mathbf{z}}) = \frac{\mathbf{z}-\mathbf{y}}{\mathbf{z}(1-\mathbf{z})}$$

$$\frac{\partial \mathbf{z}}{\partial \mathbf{h}} = \mathbf{z}(1-\mathbf{z})$$

$$\frac{\partial \mathbf{h}}{\partial \mathbf{w}} = \mathbf{X}$$

$$\frac{\partial J(\mathbf{w})}{\partial \mathbf{w}} = \mathbf{X}^T(\mathbf{z}-\mathbf{y})$$

$$\text{Gradient descent: } \mathbf{w} = \mathbf{w} - \alpha\frac{\partial J(\mathbf{w})}{\partial \mathbf{w}}$$

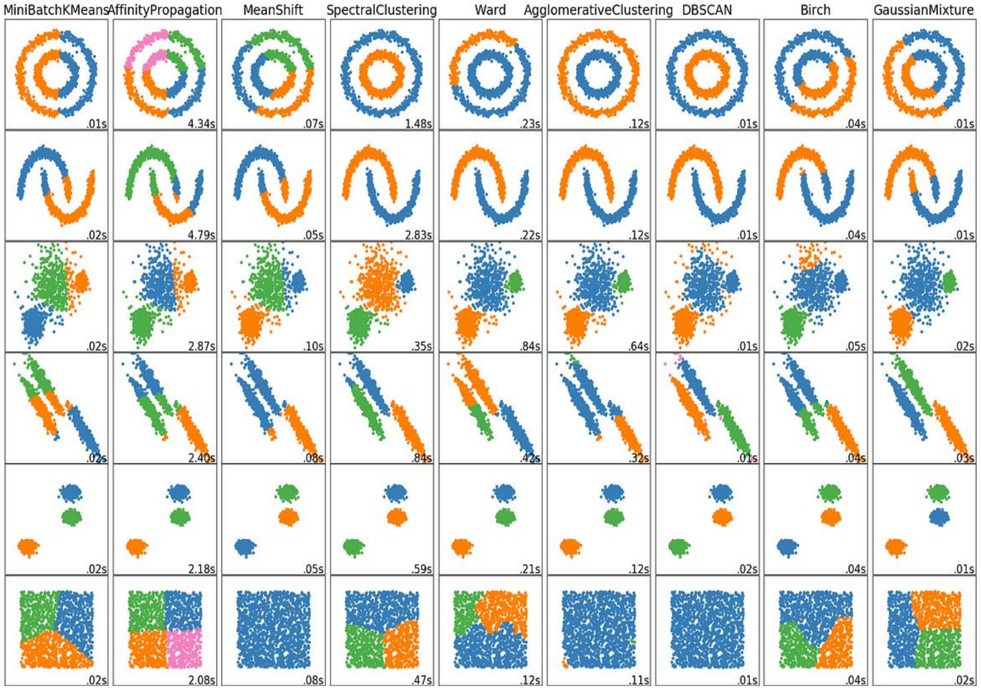
Unsupervised Machine Learning

Unsupervised learning, on the other hand, does not involve labeled data or explicit output labels. It focuses on discovering patterns, structures, or relationships within the data without any guidance from labeled examples. In the provided context, if the fishes were to learn without the explicit association of hand gestures with food availability (i.e., without labeled data), and the model had to uncover patterns or structures within the data, then it would align with unsupervised learning.

Clustering Algorithm

Clustering algorithms group similar data points together based on their characteristics or similarities. They identify natural clusters in the data without predefined labels. The algorithm measures similarity, forms clusters, and iteratively adjusts them. After convergence, each data point belongs to a specific cluster based on similarity. Clustering is used in various applications to uncover patterns and gain insights.

Clustering Algorithm



What is feature engineering?

Feature engineering is the process of preparing and transforming data to improve machine learning model performance. It involves tasks like data preprocessing, selecting relevant features, transforming features, and creating new features. It helps models better capture patterns and relationships in the data, leading to more accurate predictions.