# Basics of Machine Learning

A presentation by Manjil Mahanta

# Machine Learning?

"Field of study that gives computers the ability to learn without being explicitly programmed"

~Arthur Samuel (1959)

# Why Machine Learning?

**Machine learning has grown up as a subfield of Artificial Intelligence.**
There are only few basic things we could program a machine to do, such as how to find the shortest distance between two points, like in our gps. but for explicit and complex programs such as web search, recognize human speech, diagnose disease from X-Rays or build a self-driving car. the only way of doing it is to let a machine learn by itself.

# Types of Machine Learning

**01** Supervised Machine Learning

**02** Unsupervised Machine Learning

# Supervised Machine Learning

Machine learns the x to y or input to output mappings through examples.
The key characteristic of supervised learning is that you give your learning examples to learn from, the correct label y for a given input x, and is by seeing correct pairs of input x and desired output y the algorithm eventually learns to take just the input alone without the output label and gives a reasonably accurate prediction or guess of the output.

## Examples

**Spam Filter:** If the input x is an email and y is the output that whether the mail is spam or not spam, this supervised machine learning forms the basis of spam filter.

**Speech Recognition:** if the input is an audio clip and the algorithm's job is output the text transcript, then this is speech recognition.

**Ads:** The most money-making form of supervised learning today is probably used in online advertising. Nearly all the large online ad platforms have a learning algorithm that inputs some information about an ad and some information about you and then tries to figure out if you will click on that ad or not. Because by showing you ads they're just slightly more likely to click on, for these large online ad platforms, every click is revenue, this actually drives a lot of revenue for these companies.
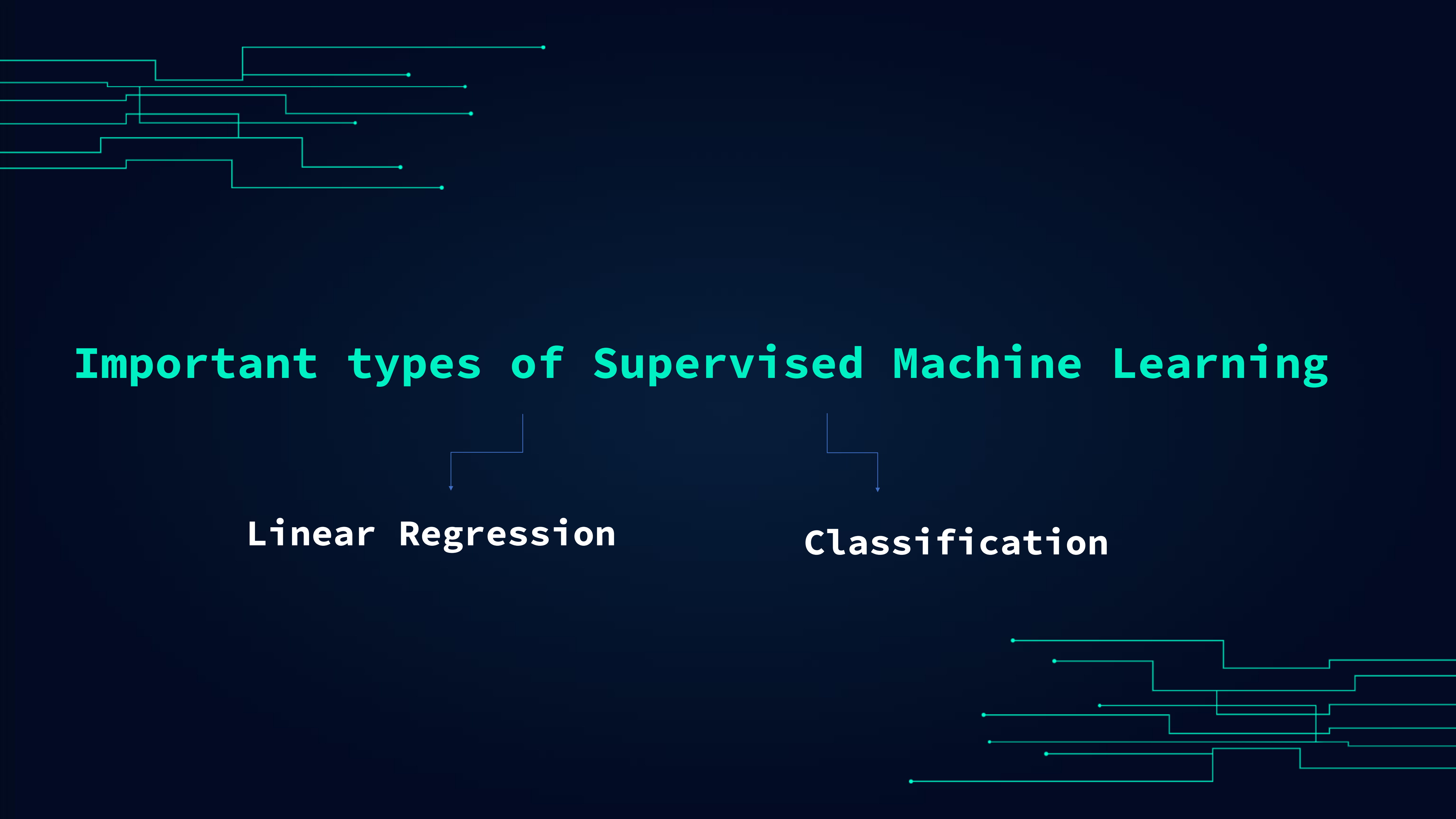
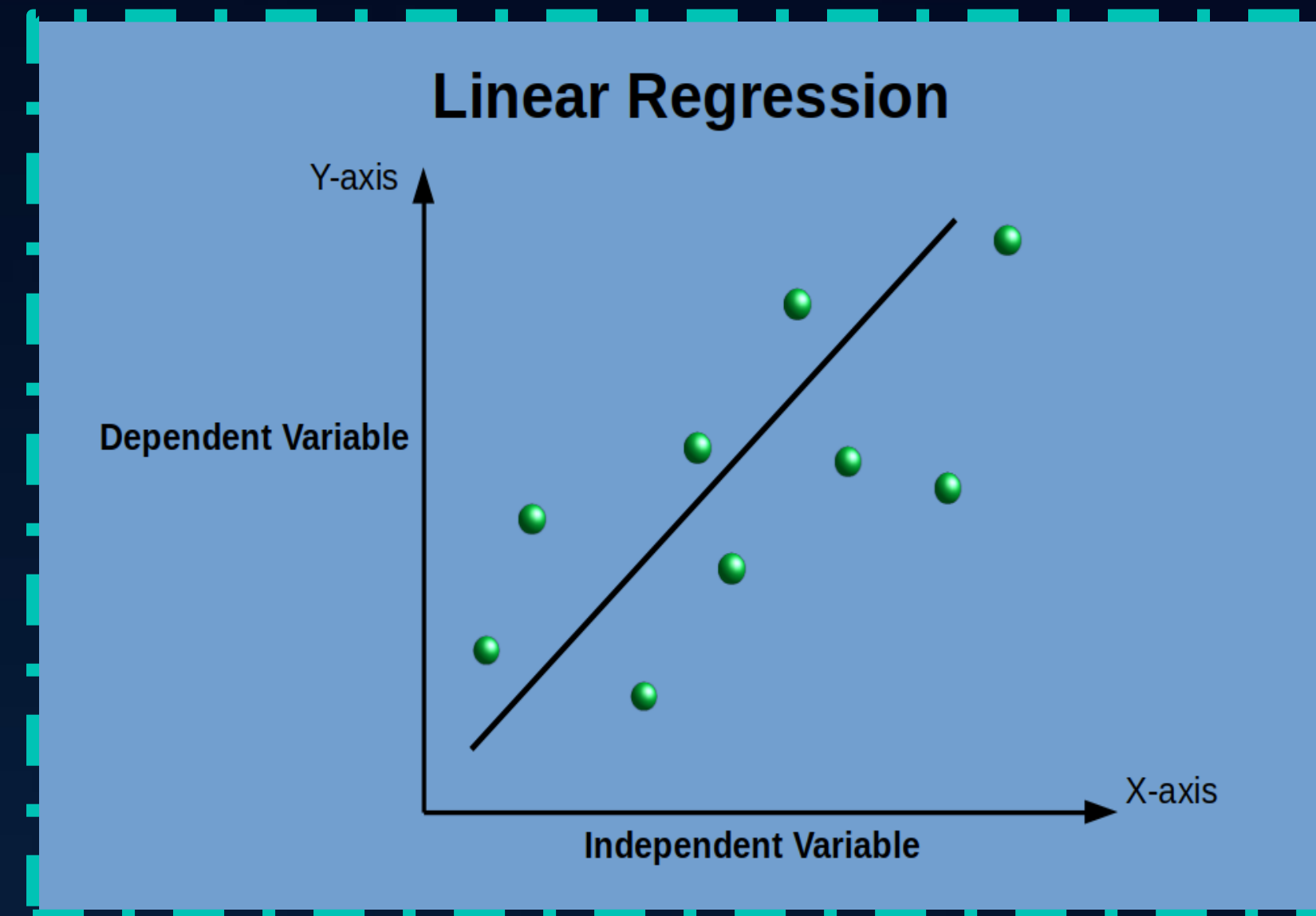# Important types of Supervised Machine Learning

Linear Regression

Classification

# Linear Regression

A statistical method that uses a straight line to model the relationship between two variables and try to predict a number from infinitely many possible numbers.



Example Of Linear Regression

Imagine you want to predict the price of houses based on their size (in square feet). You collect data on several houses, noting their sizes and prices.

Data Points:

1,000 sq ft → $150,000

1,200 sq ft → $180,000

1,500 sq ft → $220,000

1,800 sq ft → $250,000

Steps in Linear Regression:

Plot the Data: You would plot these points on a graph, with the size of the house on the x-axis and the price on the y-axis.

Fit a Line: Linear regression finds the best-fitting straight line through these points. This line represents the relationship between house size and price.

Make Predictions: Once you have the line, you can use it to predict the price of a new house. For example, if a friend has a 1,400 sq ft house, you can look at where 1,400 falls on the x-axis and see where it intersects with the line to estimate the price.

# Classification

Classification is a fundamental concept in machine learning and statistics where the goal is to teach a machine to categorize or "classify" data into predefined groups or categories. This is done by training a model on a labelled dataset, where the input data (features) is associated with the correct output (label or category). Once trained, the model can predict the category of new, unseen data based on patterns it has learned.

Example of Classification:

Breast cancer detection using machine learning is like teaching a computer to help doctors figure out if a lump in the breast is dangerous (cancerous) or not (harmless). The computer looks at data from patients, like the size of the tumor and the patient's age, and learns patterns from examples where the lumps are already labelled as either "dangerous" or "harmless." Once trained, the computer can predict whether a new lump is dangerous or not based on what it has learned. This is called classification because there are only two possible answers: dangerous or harmless. It's different from predicting numbers, which is called regression. By using this system, doctors can detect cancer earlier and save lives.
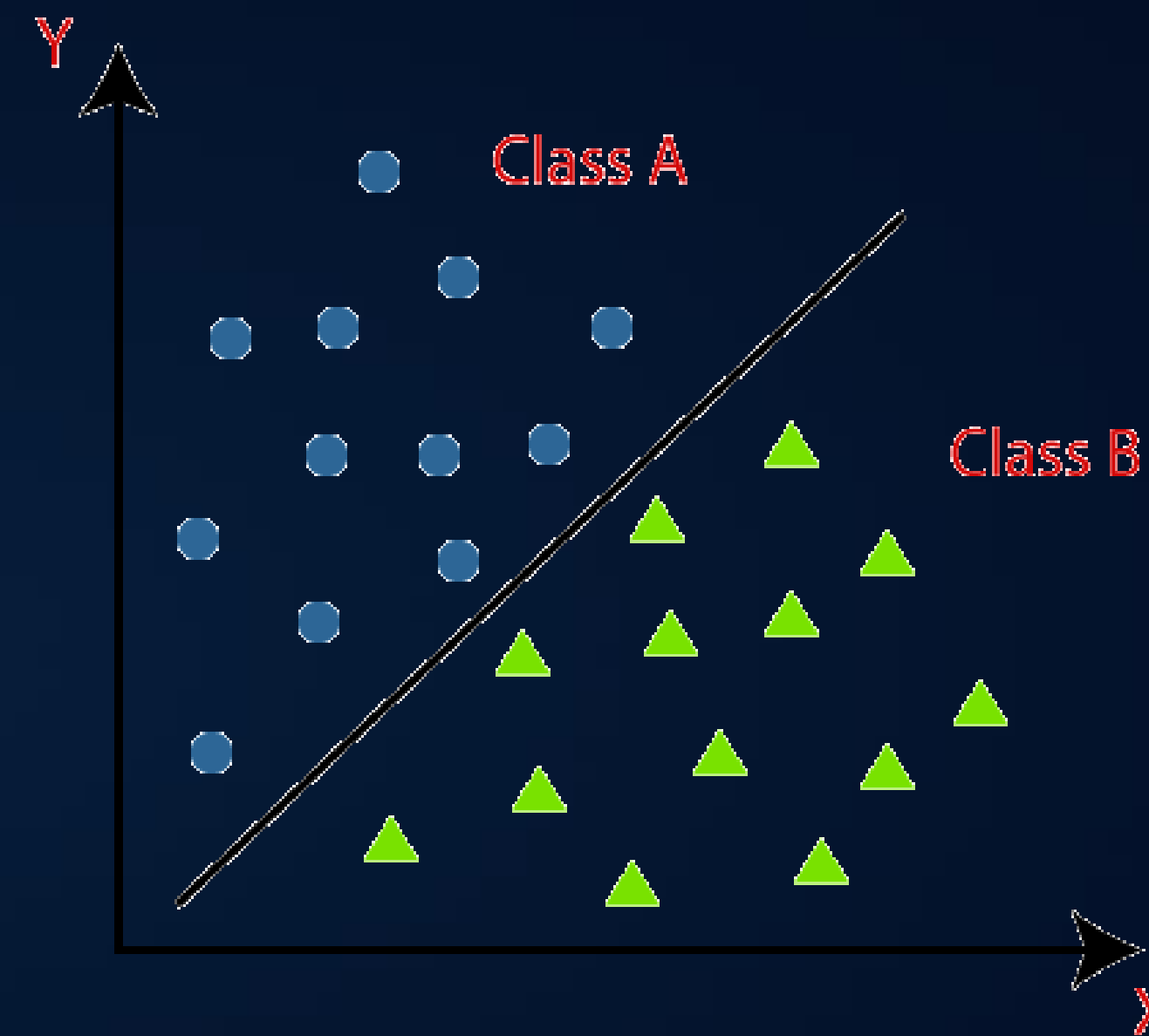
# Unsupervised Machine Learning

Unsupervised Machine learning involves analysing data that does not have associated output labels. The goal is to find patterns or structures within the data without prior guidance.

# Important Types Of Unsupervised Machine Learning

**Clustering**

**Anomaly Detection**

**Dimensionality Reduction**

## Clustering

In this type of unsupervised machine learning, algorithm groups similar data points together and forms cluster, this technique in unsupervised machine learning is called clustering.

## Anomaly Detection

Anomaly detection is a type of unsupervised learning used to identify unusual patterns or outliers in data that do not conform to expected behaviour.

## Dimensionality Reduction

Anomaly detection is a type of unsupervised learning used to identify unusual patterns or outliers in data that do not conform to expected behaviour.

# Application of Unsupervised Machine Learning

Google News uses clustering to group related news articles based on specific shared keywords.

DNA microarray data using clustering can categorise individuals based on genetic activity.

Unsupervised learning is useful in market segmentation, where companies can group customers based on behaviour or preferences without predefined categories.
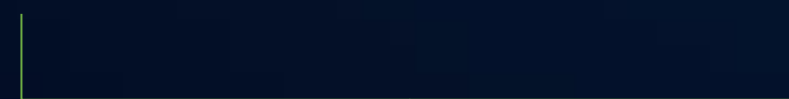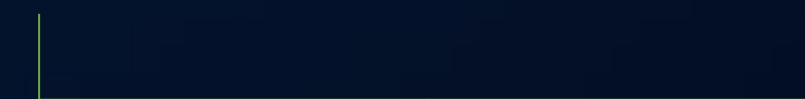
**Clustering**
v

**Fraud Detection in banking:** In a bank's transaction dataset, most transactions occur within a certain range of amounts and locations. If a customer usually makes small purchases but suddenly attempts a large transaction from a different country, this could be flagged as an anomaly. The bank's system can alert the fraud detection team to investigate further.

**Network Security:** In network monitoring, normal traffic patterns can be established based on historical data. If there is a sudden spike in data transfer or unusual access to sensitive files, this could indicate a potential security breach. Anomaly detection algorithms can help identify these irregularities, allowing for timely responses to potential threats.

**Anomaly Detection**