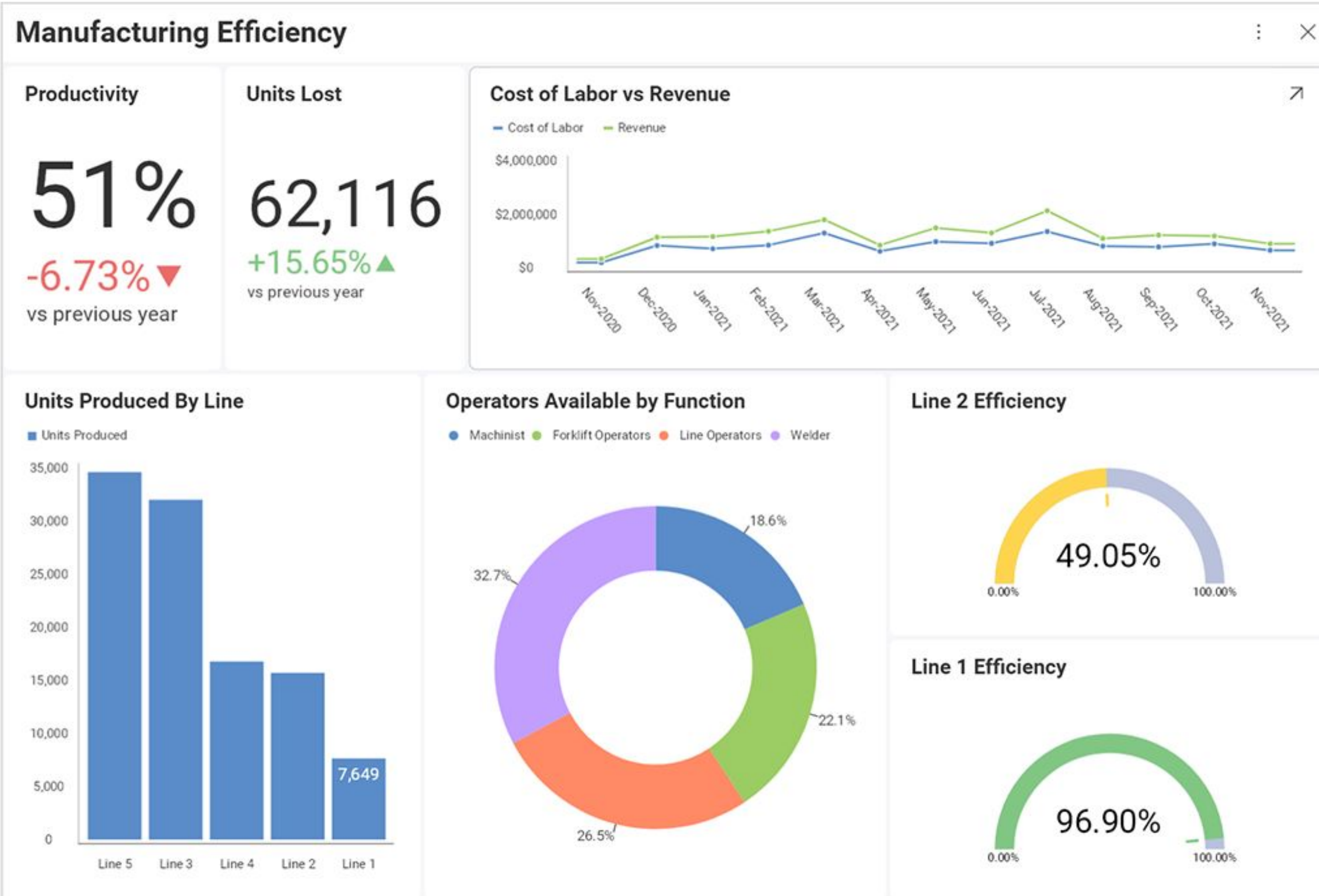
Module 1: Introduction to Data Analytics

Definition of Data Analytics

- Data Analytics is the process of examining datasets
- It extracts meaningful patterns
- Supports data-driven decision making


Manufacturing Efficiency
Productivity
51%
-6.73%
vs previous year
Units Lost
62,116
+15.65%
vs previous year
Cost of Labor vs Revenue
Cost of Labor
Revenue
$4,000,000
$2,000,000
$0
Nov-2020
Dec-2020
Jan-2021
Feb-2021
Mar-2021
Apr-2021
May-2021
Jun-2021
Jul-2021
Aug-2021
Sep-2021
Oct-2021
Nov-2021
Units Produced By Line
Units Produced
35,000
30,000
25,000
20,000
15,000
10,000
5,000
0
7,649
Line 5
Line 3
Line 4
Line 2
Line 1
Operators Available by Function
Machinist
Forklift Operators
Line Operators
Welder
18.6%
32.7%
22.1%
26.5%
Line 2 Efficiency
49.05%
0.00%
100.00%
Line 1 Efficiency
96.90%
0.00%
100.00%

Difference between Data Analysis and Data Analytics

Aspect	Data Analysis	Data Analytics
Definition	Process of inspecting, cleaning, transforming, and modeling data to find patterns	Broader process of examining data to support decision-making
Scope	Narrower	Broader
Focus	Understanding past data	Understanding past, present, and future trends
Key Question Answered	What happened? Why did it happen?	What happened? Why? What will happen? What should be done?
Nature	Technical and operational	Strategic and business-oriented
Level	Subset of data analytics	Superset (includes data analysis)
Techniques Used	Data cleaning, EDA, statistics, visualization	Data analysis, ML models, forecasting, optimization
Time Orientation	Mostly historical	Historical, real-time, and predictive
Outcome	Insights and summaries	Actionable decisions and strategies
Tools Used	Excel, SQL, Python, R	Python, R, ML tools, BI tools, Big Data platforms
Example	Finding average monthly sales	Predicting future sales and recommending actions

Data, Information, and Knowledge

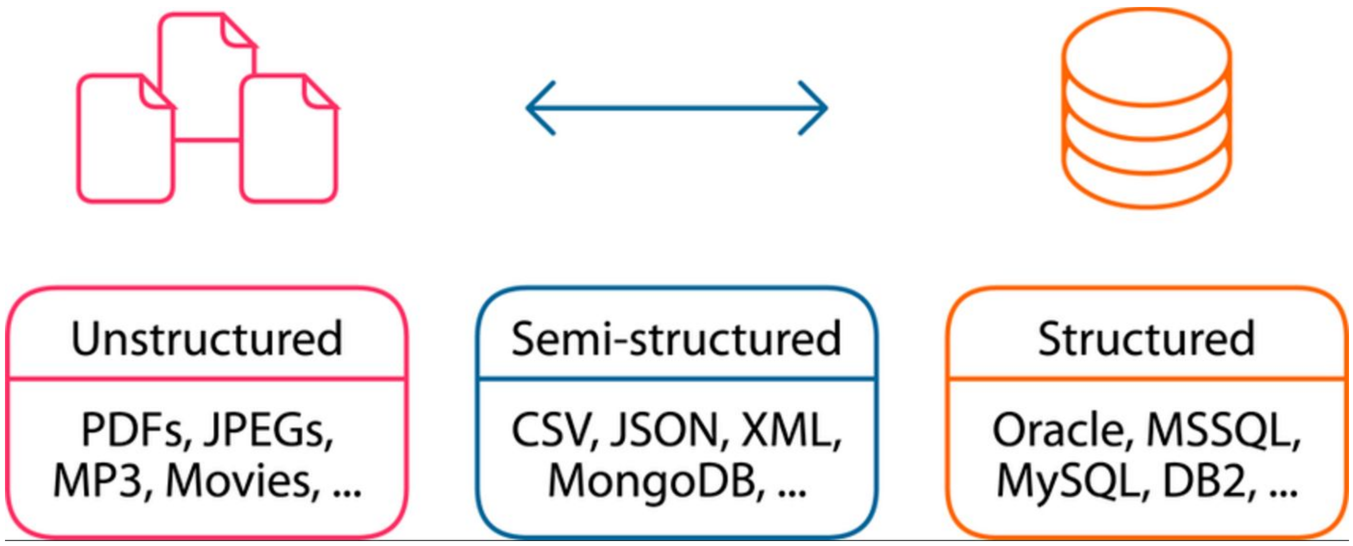
Aspect	Data	Information	Knowledge
Meaning	Raw facts	Processed data	Interpreted information
Structure	Unorganized	Organized	Contextual & experiential
Usefulness	Low	Medium	High
Role	Input	Intermediate	Output
Decision Support	No	Limited	Yes
Example	`85, 90, 78`	Average = 84.33	Student is performing well

Need for Data Analytics

- Explosion of digital data
- Better business decisions
- Automation and intelligence

Types of Data

- Structured: Tables, Databases
- Semi-structured: XML, JSON
- Unstructured: Images, Text

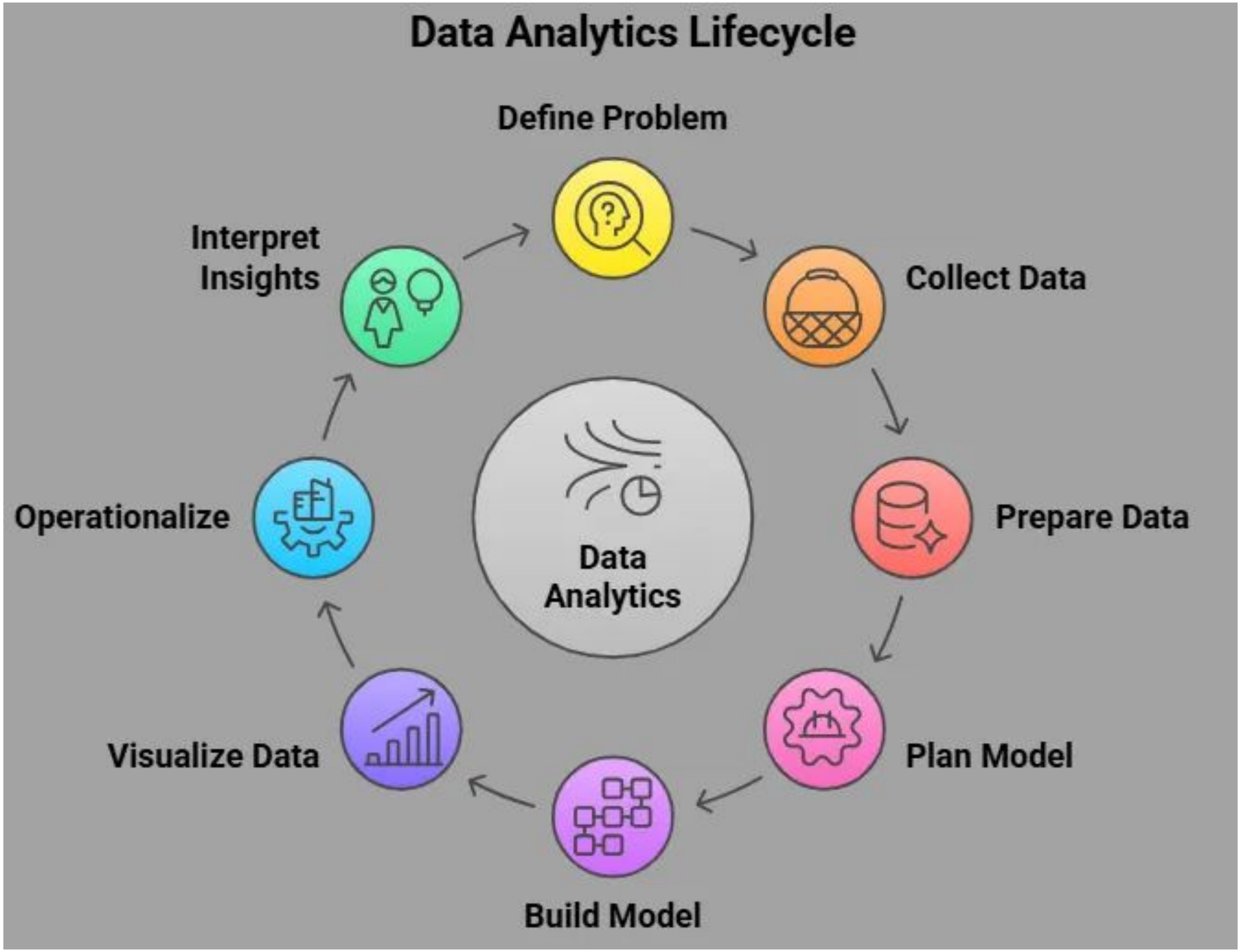


Feature	**Structured Data**	**Semi-Structured Data**	**Unstructured Data**
Organization	Data is well organized	Data is organized to some extent	Data is fully non-organized
Storage Method	Relational databases	XML, RDF, JSON	Character-based and binary data
Transaction Management	Mature transaction handling with multiple concurrency techniques	Transactions adapted from DBMS, but concurrency may cause issues	Difficult but achievable transaction management
Data Model	Tuples, rows, and tables	Tuples or graphs	No predefined data model
Schema	Schema-dependent and less flexible	More flexible than structured data	Schema-on-read
Query Capability	Complex structured queries with joins supported	Queries over anonymous nodes possible	Query performance is lowest
Flexibility	Least flexible	Moderately flexible	Most flexible
Versioning	Record-level versioning	Partial versioning	Versioning on whole data or chunks
Performance	Highest query performance	Moderate query performance	Lowest query performance
Examples	RDBMS tables, SQL databases	XML files, JSON documents	Images, videos, audio, text files

Analytics Lifecycle

- Problem Definition
- Data Collection
- Data Cleaning
- Analysis
- Visualization & Decision

✓ Understand the problem: Identify the business problem
✓ Analyze data requirements: Determine what data is needed
✓ Collect data: Gather data for use in business decision-making
✓ Prepare data: Clean data to remove errors, duplicates, or incorrect information
✓ Analyze data: Use statistical analysis to identify trends and patterns
✓ Visualize data: Use charts, graphs, and maps to represent the data
✓ Make decisions: Use insights from the data to make informed decisions


Data Analytics Lifecycle
Define Problem
Collect Data
Prepare Data
Plan Model
Build Model
Visualize Data
Operationalize
Interpret Insights
Data Analytics

Descriptive Analytics

- Summarizes historical data
- Example: Average sales per month

How It Works:

- Data Collection: Gather raw data (sales, web visits, customer reviews).
- Preparation: Clean and organize the data.
- Analysis: Apply statistical methods to find patterns and trends.
- Presentation: Display findings in easy-to-understand formats like dashboards or reports.

Numerical Example – Descriptive

- A retail company tracks sales from its stores and wants to understand performance.

Data Set (Daily Sales in $ in thousands):

Store A: {45, 50, 55, 60, 70}

Store B: {20, 40, 50, 90, 100}

Descriptive Statistics:

Mean (Average): Sum of sales / Number of days

Store A: (45+50+55+60+70) / 5 = $54,000 (Typical daily sales)

Store B: (20+40+50+90+100) / 5 = $60,000 (Higher average, but skewed)

Insight: Store B averages higher sales, but outliers (90, 100) might be misleading.

Median (Middle Value): The central point when data is ordered.

Store A (ordered): {45, 50, 55, 60, 70} = $55,000

Store B (ordered): {20, 40, 50, 90, 100} = $50,000

Insight: Store A's median ($55k) is close to its mean ($54k), showing fairly consistent sales; Store B's median ($50k) is lower than its mean ($60k), highlighting the impact of high-performing days.

Mode (Most Frequent):

Store A: No single mode (or all are modes)

Store B: No single mode

Insight: Useful for identifying popular price points or product sales.

Range (Variability): Max value - Min value

Store A: $70 - $45 = $25,000 (Lower variability)

Store B: $100 - $20 = $80,000 (Higher variability)

Insight: Store B's sales are much less predictable and consistent than Store A's.

Conclusion:

Descriptive analytics summarizes the data to reveal that while Store B has a higher average, Store A has more consistent performance, helping managers decide where to focus efforts (e.g., promotions in Store A to boost consistency, or investigating outliers in Store B).

Numerical Example – Descriptive Cont..

Month	Advertising Spend (X_1)	Sales (X_2)	Footfall (X_3)
Jan	10	120	8
Feb	12	135	9
Mar	11	128	8.5
Apr	15	160	11
May	14	155	10.5
Jun	13	145	9.5

Using **descriptive analytics**, compute:

- Mean of each variable
- Standard Deviation of Sales
- Covariance between Advertising Spend and Sales
- Correlation between Advertising Spend and Sales
- Interpret the results

Graph-based Descriptive Analytics

- It uses nodes (entities) and edges (relationships) to visualize and summarize complex historical data, revealing patterns like central users, clusters, or paths, complementing traditional charts (histograms, scatter plots) by focusing on *connections* to show *how* things are linked

Nodes: Represent entities like people, products, or locations.

Edges: Represent relationships between nodes, with properties like direction (e.g., "follows") and strength (e.g., "frequently calls").

Graph Analytics: Analyzes these nodes and edges to uncover hidden structures, centralities, and clusters in the data.

Graph-based Descriptive Analytics cont..

Descriptive Insights from Graphs

- Centrality: Identifying the most important nodes (e.g., most connected users).
- Path Analysis: Finding shortest or most common routes between entities (e.g., degrees of separation).
- Community Detection: Grouping densely connected nodes (clusters).
- Network Density: Understanding how interconnected a network is overall.

Numerical Example – Descriptive Cont..

Month	Sales
Jan	120
Feb	135
Mar	128
Apr	160
May	155
Jun	145
Jul	150
Aug	158
Sep	165
Oct	170
Nov	168
Dec	180

Using **graph-based descriptive analytics**:
- Draw a **Line Graph**
- Draw a **Bar Chart**
- Identify **trend, peak, and variability**
- Provide interpretation

Interpretation (Very Important for Exams)

- Line graph indicates a **positive sales trend**
- Bar chart highlights **December as peak month**
- Business performance improves steadily across the year
- Graphs help in **quick decision-making** compared to raw data

Predictive Analytics

It is a branch of advanced analytics that uses historical and current data, statistical algorithms, and machine learning (ML) techniques to identify patterns and forecast future outcomes.

- Uses models to predict future outcomes

Predictive Analytics Cont..

- **Regression Analysis:** Estimates relationships between variables to predict a numerical value (e.g., forecasting next quarter's sales or a property's price).
- **Classification Models:** Categorizes data into predefined groups to answer binary "yes/no" questions (e.g., "Is this transaction fraudulent?" or "Will this customer do fraud?").
- **Clustering Models:** Groups similar data points based on shared characteristics without predefined labels, often used for customer segmentation.
- **Time Series Models:** Specifically designed for data points captured over time to detect seasonal trends or cyclical patterns (e.g., predicting energy demand peaks).
- **Neural Networks:** Complex AI algorithms modeled after the human brain that excel at finding nonlinear relationships in vast, unstructured datasets like images or voice recordings

Regression Analysis

- It is a statistical method used to understand the relationship between input features and a target value that varies across a continuous numeric range.
- helps measure how changes in different factors affect the outcome, allowing better predictions, planning and decision-making across various fields.

Regression Analysis cont..

Types of Regression:

- Simple Linear Regression
- Multiple Linear Regression
- Polynomial Regression
- Logistic Regression (classification)

Simple Linear Regression

- Models relationship between one independent variable X and dependent variable Y.
- Equation: $Y = a + bX + \varepsilon$

where

a: Intercept

b: Slope

ε: Error term

X: Independent variable

Y: Dependent variable

Simple Linear Regression cont..

Objectives:

- Minimize prediction error.
- Estimate best fitting line using Least Squares Method.

Simple Linear Regression cont..

Example:

Dataset

Advertising Spend (X)	Sales (Y)
10	120
12	135
15	160
18	175
20	190

$$b = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - (\sum X)^2}$$

$$a = \frac{\sum Y - b\sum X}{n}$$

Example cont..

- Interpretation

Slope (6.91) i.e., b:
For every ₹1 lakh increase in advertising spend, sales increase by approximately ₹6.91 lakhs.

Intercept (52.35) i.e., a:
If advertising spend is zero, expected sales are ₹52.35 lakhs (baseline sales).

- **Prediction Example**

Predict sales when **Advertising Spend = 16**

$$Y = 52.35 + 6.91(16) = 52.35 + 110.56 = 162.91$$

Example 2

Student Roll No.	Study Hours (X)	Score (Y)
1	2	50
2	4	55
3	5	65
4	7	70
5	8	78
6	10	85
7	12	88

- Obtain the regression equation of Y on X
- Predict the score for X = 9 hours
- Compute residuals
- Calculate SSE (Sum of Squared Errors)
- Interpret the error term

Simple Linear Regression cont..

What is R^2 (Coefficient of Determination)?

- R^2 measures the proportion of variance in the dependent variable that is explained by the independent variable(s) in a regression model.

$$R^2 = 1 - \frac{SSE}{SST}$$

Multiple Linear Regression

- It predicts a single dependent variable (outcome) using multiple independent variables (predictors) to find a linear relationship, extending simple regression by handling complex scenarios.

$Y = a + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon$

Y = Dependent variable (target)

$X_1, X_2 \ldots X_k$ = Independent variables (features)

a = Intercept

$\beta_1, \beta_2, \ldots \beta_k$ = Regression coefficients

ε = Error term

Multiple Linear Regression Numerical Example

- Advertising Spend (X_1)
- Number of Salespersons (X_2)

Observation	X_1	X_2	Y
1	1	1	6
2	2	1	8
3	1	2	9
4	2	2	11

- Compute Final Regression Equation
- Insight

Normal Equations for coefficients:

$$\sum Y = na + b_1 \sum X_1 + b_2 \sum X_2$$

$$\sum X_1 Y = a \sum X_1 + b_1 \sum X_1^2 + b_2 \sum X_1 X_2$$

$$\sum X_2 Y = a \sum X_2 + b_1 \sum X_1 X_2 + b_2 \sum X_2^2$$

- **Final Regression Equation**

$$\boxed{Y = 1 + 2X_1 + 3X_2}$$

❖ ***Insight:***

Intercept (a = 1)

- Base sales when both predictors are zero
- Represents fixed or minimum sales.

Coefficient of X_1 ($b_1 = 2$)

- Keeping X_2 constant, **1-unit increase in advertising** increases sales by **2 units**

Coefficient of X_2 ($b_2 = 3$)

- Keeping X_1 constant, **1-unit increase in salespersons** increases sales by **3 units**

Polynomial Regression

- It is a type of regression that models the relationship between an independent variable (x) and a dependent variable (y) as an n-th degree polynomial, allowing it to capture non-linear patterns.

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \ldots + \beta_n x^n + \epsilon$$

- y is the dependent variable.
- x is the independent variable.
- $\beta_0, \beta_1, \ldots, \beta_n$ are the coefficients of the polynomial terms.
- n is the degree of the polynomial.
- ϵ represents the error term.

Polynomial Regression cont..

- Numerical Example:

X (Hours)	Y (Marks)
1	52
2	60
3	75
4	92

- Fit a second-degree polynomial regression model.
- Polynomial Model

$$Y = a + bX + cX^2$$

- Normal Equations for coefficients:

$$\sum Y = na + b\sum X + c\sum X^2$$

$$\sum XY = a\sum X + b\sum X^2 + c\sum X^3$$

$$\sum X^2Y = a\sum X^2 + b\sum X^3 + c\sum X^4$$

- Compute Final Regression Equation
- Interpretation

Polynomial Regression cont..

- **Final Polynomial Regression Equation**

$$Y = 45 + 2X + 5X^2$$

Insight:

Intercept (a = 45)

- Baseline marks when study hours are zero

Linear Term (b = 2)

- Initial increase in marks with study hours

Quadratic Term (c = 5)

- Marks increase **faster** as study hours increase
- Captures the **curved (non-linear) trend**

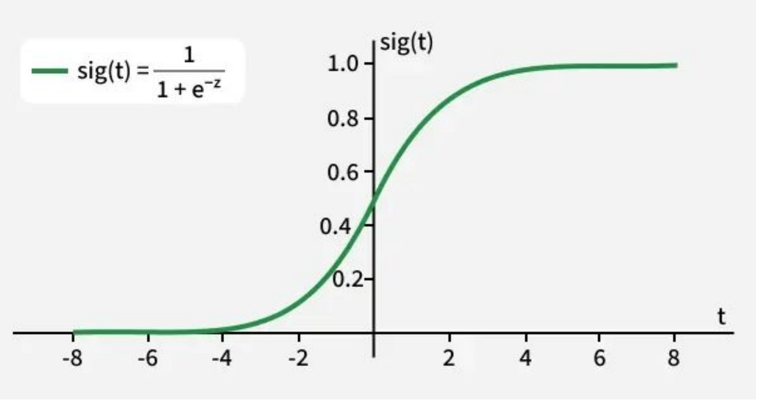
Logistic Regression

- Is a powerful supervised learning method for predicting categorical outcomes (like Yes/No, Spam/Not Spam, Buy/Not Buy) by modeling the probability of an event using the sigmoid (logistic) function, which transforms linear outputs into probabilities between 0 and 1

$$P(Y=1) = \frac{1}{1+e^{-z}}$$

If $P(Y=1) \geq 0.5 \Rightarrow$ Class 1

Else $\Rightarrow$Class 0



Logistic Regression cont..

- Example:

Predict whether a student passes (1) or fails (0) based on study hours (X).

Model: $z = -4 + 1.2X$

z= −4 + 1.2(4) = 0.8

Apply sigmoid:

$$P = \frac{1}{1 + e^{-0.8}} \approx 0.69$$

Interpretation

- Probability of passing = **69%**
- Since $P > 0.5$, student is predicted to **pass**

Logistic Regression cont..

Example: A company wants to predict whether a customer will **buy a product (1)** or **not buy (0)** based on **monthly income (X)** in lakhs.

Customer	**Income X (lakhs)**	**Purchase Y**
1	1	0
2	2	0
3	3	0
4	4	1
5	5	1

- Logistic regression uses iterative optimization: Assume initial values: $a=-4$, $b=1$
- Model for one variable: z=a + bx
- Assume loss is minimum at a=−6,b=2

Logistic Regression cont..

- Final Logistic Regression Model:

$$\boxed{z = -6 + 2X}$$

$$\boxed{P(Y = 1) = \frac{1}{1 + e^{-(-6+2X)}}}$$

- Prediction Using Final Model

  For X = 4

$$z = -6 + 2(4) = 2$$

$$P = \frac{1}{1 + e^{-2}} = 0.881$$

- Interpretation of Coefficients

Intercept (a = −6)

- Base tendency when income is zero
- Shifts decision boundary

Coefficient (b = 2)

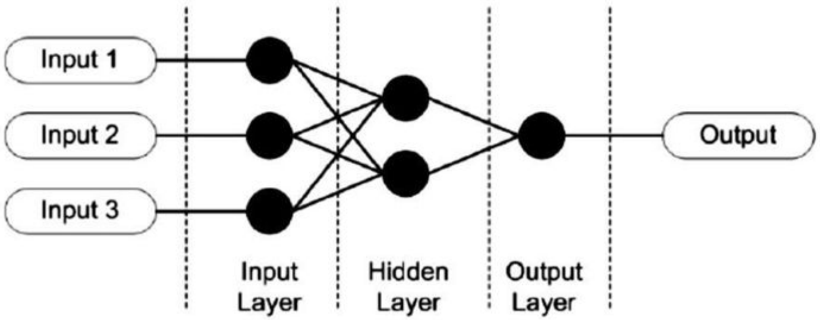
$$\text{Odds Ratio} = e^{2} = 7.3$$

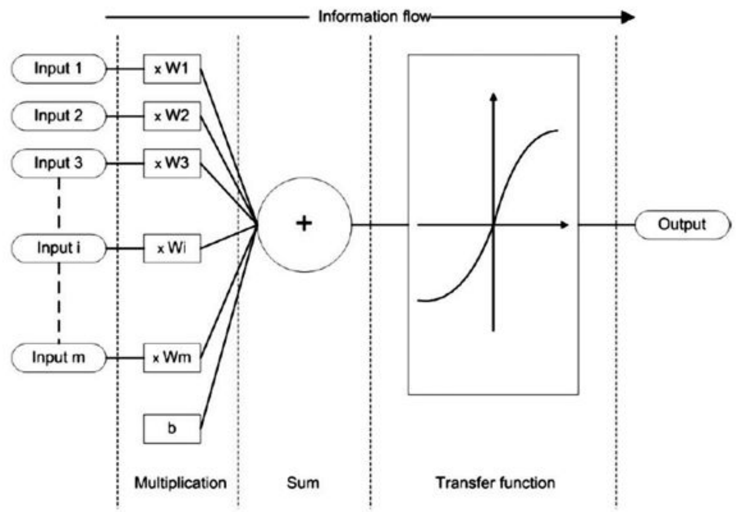
Every 1 lakh increase in income multiplies buying odds by 7.3

Neural Networks

- The Neural networks leverages brain-inspired models to find complex, non-linear patterns in large datasets for forecasting future outcomes
- These models use layers of interconnected nodes (neurons) to process input data, adjusting weights through backpropagation to improve accuracy in tasks
- The model has three simple sets of rules: multiplication, summation and activation

Neural Networks cont..





- Input Layer: Receives historical data (e.g., customer behavior, sales figures).
- Hidden Layers: Nodes (neurons) apply complex functions to transform inputs, learning intricate patterns not obvious in simple rules.
- Output Layer: Produces the final prediction (e.g., next quarter's sales, likelihood of a customer leaving).

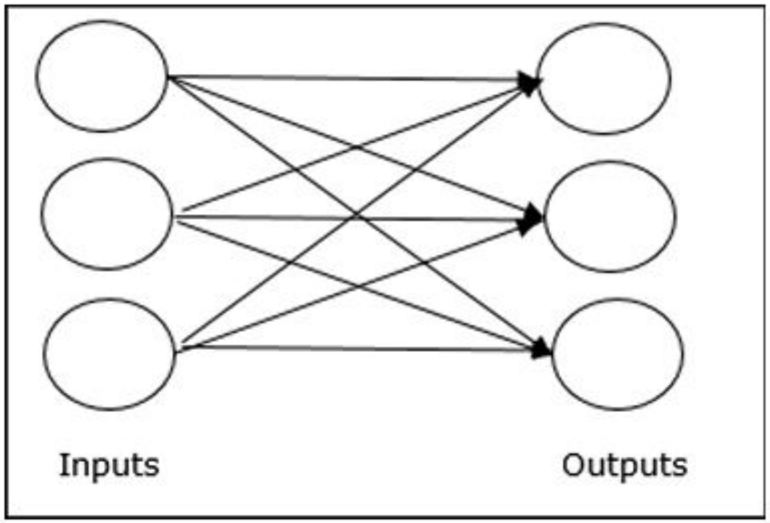
Neural Networks cont..

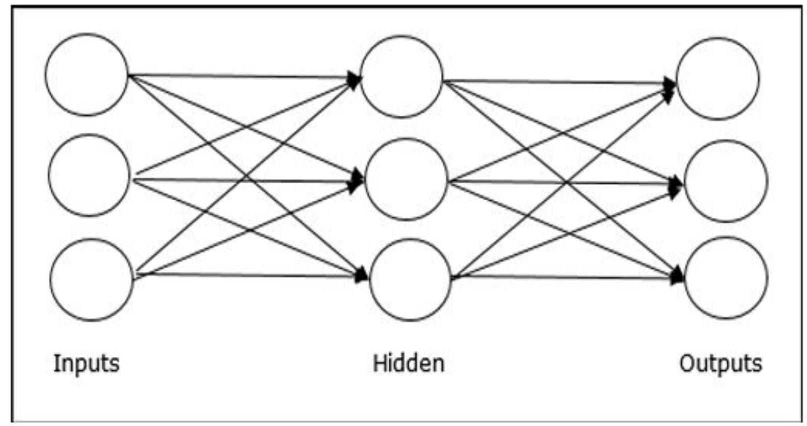
- Basic building blocks of Neural Networks:

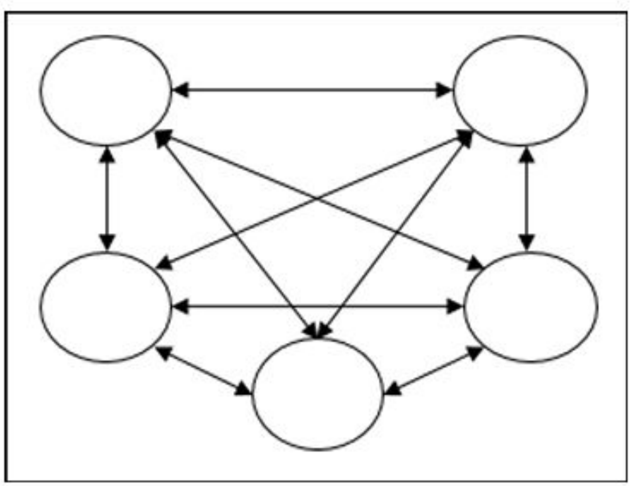
1. Network Topology
2. Adjustments of Weights or Learning
3. Activation Functions

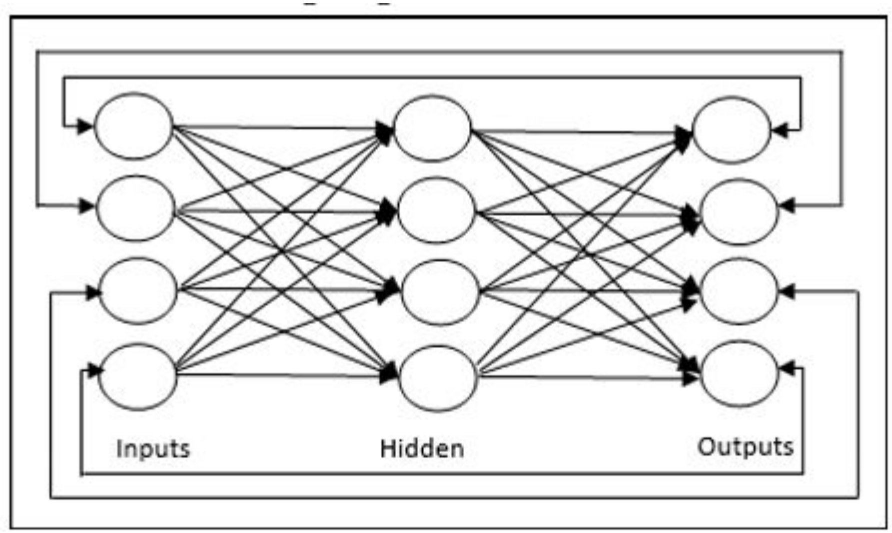
Neural Networks cont..

- Network Topology (feed forward, and Recurrent Neural Networks)





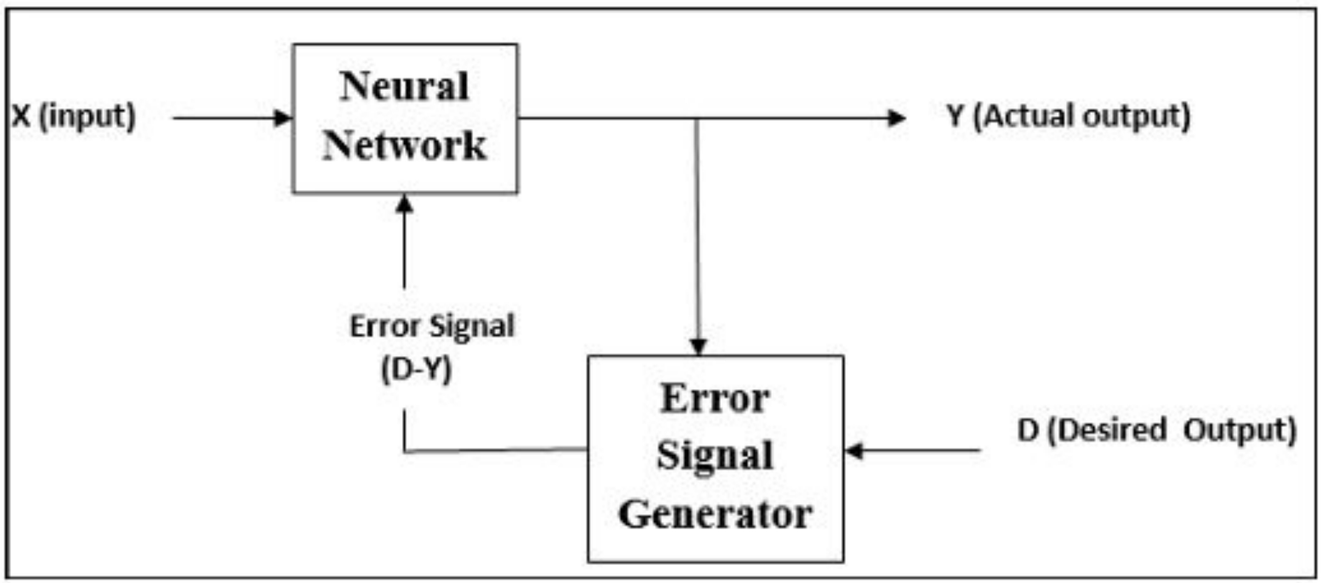




Neural Networks cont..

- Adjustments of Weights or Learning:

Learning, in artificial neural network, is the method of modifying the weights of connections between the neurons of a specified network.



Neural Networks cont..

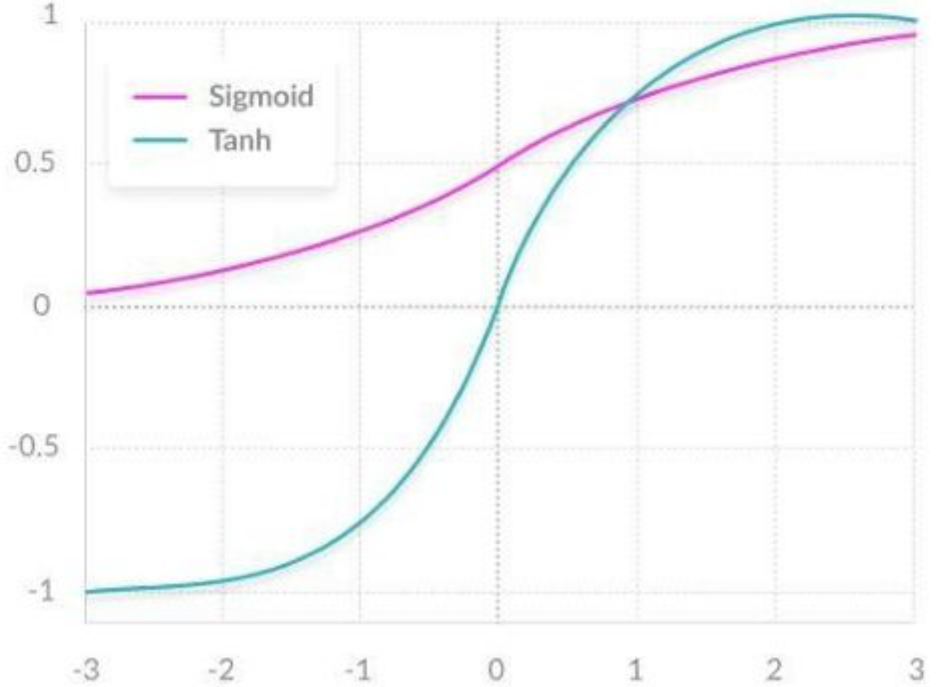
- **Activation Functions** : An activation function is a mathematical equation that determines the output of each element (perceptron or neuron) in the neural network.

1. The sigmoid function has a smooth gradient and outputs values between zero and one. For very high or low values of the input parameters, the network can be very slow to reach a prediction, called the vanishing gradient problem.
2. The TanH function is zero-centered making it easier to model inputs that are strongly negative strongly positive or neutral.
3. The ReLu function is highly computationally efficient but is not able to process inputs that approach zero or negative.
4. The Leaky ReLu function has a small positive slope in its negative area, enabling it to process zero or negative values.
5. The Parametric ReLu function allows the negative slope to be learned, performing backpropagation to learn the most effective slope for zero and negative input values.
6. Softmax is a special activation function use for output neurons. It normalizes outputs for each class between 0 and 1, and returns the probability that the input belongs to a specific class.
7. Swish is a new activation function discovered by Google researchers. It performs better than ReLu with a similar level of computational efficiency.

Neural Networks cont..

Activation Function	Mathematical Expression	Output Range
Sigmoid	$\sigma(x) = \frac{1}{1+e^{-x}}$	$(0,\ 1)$
Tanh	$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$	$(-1,\ 1)$
ReLU	$f(x) = \max(0, x)$	$[0,\ \infty)$
Leaky ReLU	$f(x) = \max(\alpha x, x),\ \alpha \ll 1$	$(-\infty,\ \infty)$
Parametric ReLU (PReLU)	$f(x) = \max(ax, x),\ a \text{ learned}$	$(-\infty,\ \infty)$
Softmax	$f_i(x) = \frac{e^{x_i}}{\sum_j e^{x_j}}$	$(0,\ 1)$, sum = 1
Swish	$f(x) = x \cdot \sigma(x)$	$(-\infty,\ \infty)$

Neural Networks cont..



Two common neural network activation functions - Sigmoid and Tanh

Neural Networks cont..

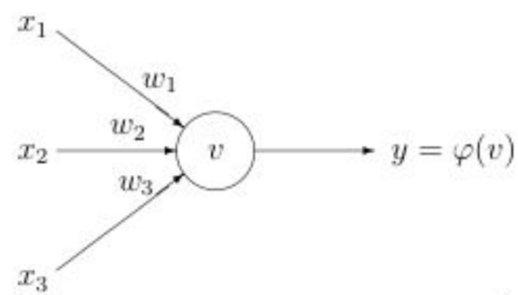
- Numerical Example:

Consider the unit shown in figure below. Suppose that the weights corresponding to the three inputs have the following values:

W1= 2

W2= -4

W3= 1



and the activation of the unit is given by the step-function:

$$\varphi(v) = \begin{cases} 1 & \text{if } v \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Calculate what will be the output value y of the unit for each of the following input patterns:

Pattern	P_1	P_2	P_3	P_4
x_1	1	0	1	1
x_2	0	1	0	1
x_3	0	1	1	1

Neural Networks cont..

- Numerical Example 1:

Feature	Value
Study Hours (X_1)	4
Attendance (X_2)	70
Actual Output (Y)	1

Initial Weights and Biases

Hidden Layer

$w_1 = 0.5, w_2 = 0.01, b_h = -2$

Output Layer

$w_3 = 1.2, b_o = -1$

Input layer: 2 neurons (X1 ,X2)

Hidden layer: 1 neuron

Output layer: 1 neuron

Activation: Sigmoid

Neural Networks cont..

- Numerical Example 2:

Feature	Value
House Size (X_1)	6
Bedrooms (X_2)	3
Actual Output (Y)	1 (High price)

- Predict House Price (Low = 0, High = 1) based on:
- X_1 = House Size (in 100 sq.ft units)
- X_2 = Number of Bedrooms

Neural Network Architecture
Input layer: 2 neurons (X_1, X_2)
Hidden layer: 1 neuron
Output layer: 1 neuron
Activation function: Sigmoid

Input → Hidden

$$w_1 = 0.5, w_2 = 0.3, b_h = -1$$

Hidden → Output

$$w_3 = 1.2, b_o = -0.5$$

Neural Networks cont..

- Numerical Example 3:

Predict whether a customer will purchase a product (Yes = 1, No = 0) based on:

1. X_1 = Monthly Income (in ₹10,000 units)
2. X_2 = Time Spent on Website (minutes)
3. X_3 = Number of Previous Purchases

Feature	Value
X_1 (Income)	5
X_2 (Time)	10
X_3 (Purchases)	2
Actual Output (Y)	1

Neural Network Architecture
Input layer: 3 neurons (X_1, X_2, X_3)
Hidden layer: 2 neurons
Output layer: 1 neuron
Activation function: Sigmoid (all neurons)

Input → Hidden Layer
Hidden neuron **H_1**

$$w_{11} = 0.4,\ w_{12} = 0.3,\ w_{13} = 0.2,\ b_1 = -2$$

Hidden neuron **H_2**

$$w_{21} = 0.1,\ w_{22} = 0.6,\ w_{23} = 0.5,\ b_2 = -3$$

Hidden → Output Layer

$$v_1 = 1.2, v_2 = 0.8, b_o = -1$$

Neural Networks cont..

- **Numerical Example 4 (Backpropagation):**

Predict whether a student passes (1) or fails (0) based on; X = Study Hours

Study Hours (X)	Actual Output (Y)
4	1

Neural Network Structure

Input layer: 1 neuron

Hidden layer: 1 neuron

Output layer: 1 neuron

Activation function: Sigmoid

Loss function: Squared Error

Given Information:

Input → Hidden

$$w_1 = 0.5, b_1 = 0$$

Hidden → Output

$$w_2 = 1.0, b_2 = 0$$

Learning rate

$$\eta = 0.1$$

Note: Formulas for backpropagation and weight update:

$$\delta_2 = (\hat{Y} - Y)\hat{Y}(1 - \hat{Y})$$

$$w_2^{new} = w_2 - \eta(\delta_2 a_1)$$

$$\delta_1 = \delta_2 w_2 a_1(1 - a_1)$$

$$w_1^{new} = w_1 - \eta(\delta_1 X)$$

Neural Networks cont..

- **Numerical Example 5 (Backpropagation):**

Predict whether a student passes (1) or fails (0) based on; X = Study Hours

Data Point	Study Hours (X)	Result (Y)
1	2	0
2	4	1

Neural Network Structure

Input layer: 1 neuron

Hidden layer: 1 neuron

Output layer: 1 neuron

Activation function: Sigmoid

Loss function: Mean Squared Error

Learning rate: $\eta = 0.1$

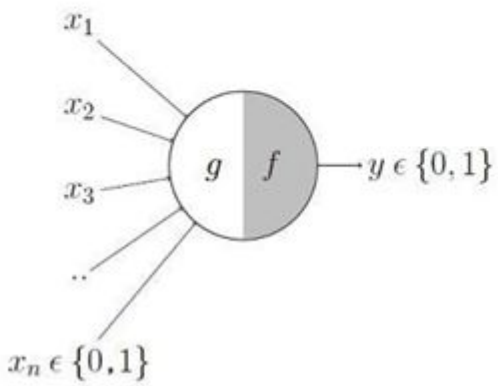
Given Information:

Initial Weights and Biases

$$w_1 = 0.5, w_2 = 0.8$$
$$b_1 = 0, b_2 = 0$$

MCCULLOGH-PITTS MODEL

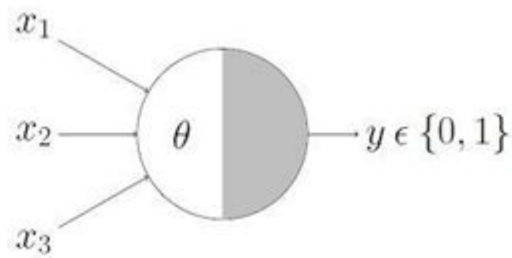
- In 1943 two electrical engineers, Warren McCullogh and Walter Pitts, published the first paper describing what we would call a neural network.
- It may be divided into 2 parts. The first part, g takes an input, performs an aggregation and based on the aggregated value the second part, f makes the decision.



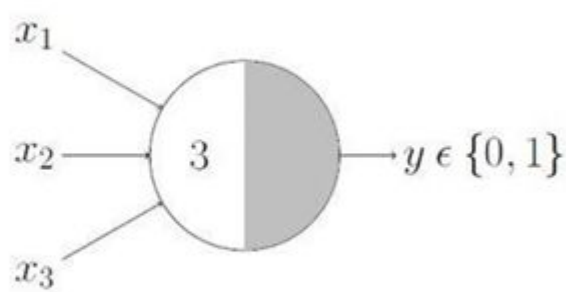
- The McCulloch-Pitts neural model is also known as linear threshold gate. It is a neuron of a set of inputs I1, I2,I3,…Im and one output ‘y’ . The linear threshold gate simply classifies the set of inputs into two different classes. Thus, the output y is binary.

MCCULLOGH-PITTS MODEL cont..

- Boolean functions using Mcculogh-Pitts model: all inputs are Boolean and the output is also Boolean.



- AND function: An AND function neuron would only fire when ALL the inputs are ON i.e., g(x) ≥ 3 here.



MCCULLOGH-PITTS MODEL cont..

- Draw for -

OR function, NOT function, NOR function

Classification Models

- Classification models in predictive analytics categorize data into predefined groups (classes) to forecast outcomes.
- Types of Classification:

Binary, Multiclass, Multilabel.

- Models: Decision Tree, Neural Networks, SVM, Random Forest, Logistic Regression.
- Application Areas: Finance, Healthcare, Marketing, Image Recognition etc.

Classification Models cont..

Evaluation Metrics:

Metric	Meaning
Accuracy	Correct predictions
Precision	Quality of positive predictions
Recall	Coverage of actual positives
F1-score	Balance of precision & recall
Confusion Matrix	Prediction breakdown

Time Series Models

- Time series models in predictive analytics analyze sequential, time-stamped data to forecast future values by identifying patterns like trends and seasonality in historical data
- Methods used: ARIMA, Exponential Smoothing, and advanced ML/DL (like LSTMs)
- Key components of time series:

Component	Meaning	Example
Trend	Long-term movement	Increasing sales
Seasonality	Repeating pattern	Higher sales in festivals
Cyclic	Irregular cycles	Economic cycles
Noise	Random variation	Measurement error

Time Series Models cont..

- Time Series Models:

Naïve Forecasting- Assumes next value equals last value

$$\hat{Y}_{t+1} = Y_t$$

Moving Average (MA) Averages last $k$observations

$$\hat{Y}_t = \frac{1}{k}\sum_{i=1}^{k} Y_{t-i}$$

Simple Exponential Smoothing (SES) Gives more weight to recent observations, Where $0 < \alpha < 1$

$$\hat{Y}_t = \alpha Y_{t-1} + (1-\alpha)\hat{Y}_{t-1}$$

Time Series Models cont..

- Numerical Example 1:

Month	Actual Sales
Jan	200
Feb	220
Mar	210
Apr	230
May	240

Forecast **next month's sales** using historical monthly sales data.

Model Used: 2-Month Moving Average (MA)

Time Series Models cont..

- Numerical Example 2:

Month	Power Demand (MW)
Jan	820
Feb	780
Mar	800
Apr	860
May	900
Jun	940
Jul	910
Aug	970

An electricity board records **monthly power demand (in MW)**.
Using a **3-month Moving Average**:

- Forecast demand
- Compute **Mean Absolute Error (MAE)**
- Interpret the result

Time Series Models cont..

- Numerical Example 3: Time Series Forecasting using Exponential Smoothing

Month	Power Demand (MW)
Jan	820
Feb	780
Mar	800
Apr	860
May	900
Jun	940
Jul	910
Aug	970

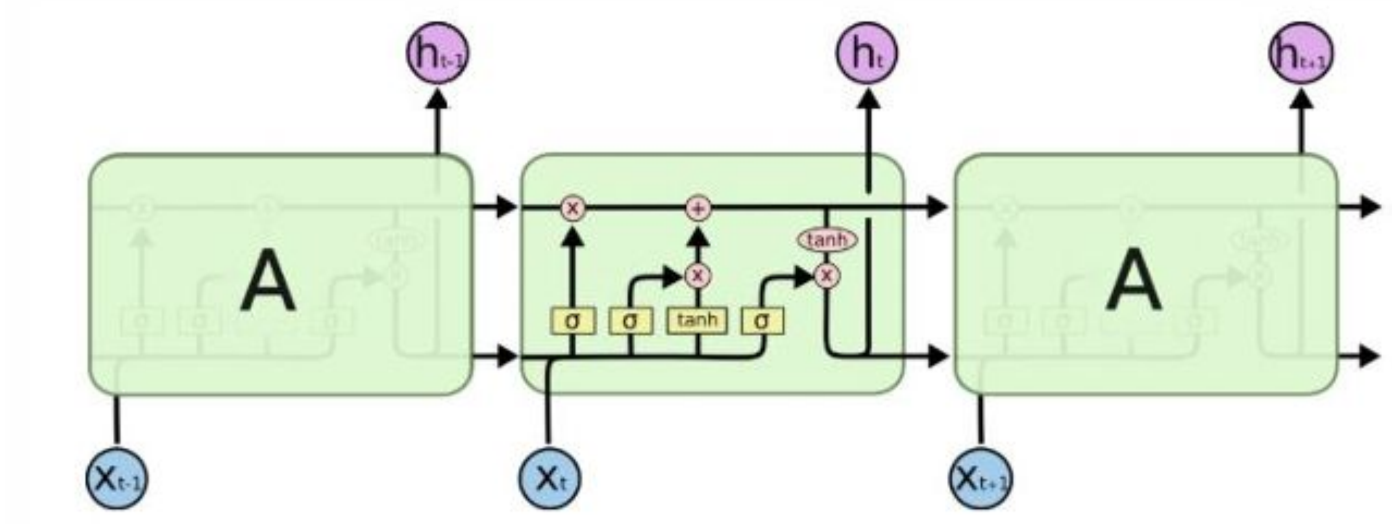
$$\hat{Y}_t = \alpha Y_{t-1} + (1-\alpha)\hat{Y}_{t-1}$$

Where, $\alpha=0.3$

- Forecast demand for each month using Simple Exponential Smoothing (SES)
- Compute **Mean Absolute Error (MAE)**

LSTM Cell Unfolded

- LSTM combines gated memory and nonlinear transformations to capture long-term dependencies in time series data.



$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f)$$

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i)$$

$$\tilde{C}_t = \tanh(W_c[h_{t-1}, x_t] + b_c)$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o)$$

$$h_t = o_t \cdot \tanh(C_t)$$

LSTM Cell cont..

- Numerical 1: Forecast the **next value** in a small time series using **one LSTM cell**

Time (t)	Value (x□)
1	2
2	4
3	6

Assumptions:

- Single input feature
- Single LSTM cell
- Initial hidden state:

$$h_0 = 0$$

- Initial cell state:

$$C_0 = 0$$

Weights & Biases:

(All weights = 0.5, all biases = 0)

Equations used:

$$f_t = \sigma(W_f(h_{t-1} + x_t))$$

$$i_t = \sigma(W_i(h_{t-1} + x_t))$$

$$\tilde{C}_t = \tanh(W_c(h_{t-1} + x_t))$$

$$C_t = f_t C_{t-1} + i_t \tilde{C}_t$$

$$o_t = \sigma(W_o(h_{t-1} + x_t))$$

$$h_t = o_t \tanh(C_t)$$

LSTM Cell cont..

- Numerical 2: Forecast the **next value** in a time series using **one LSTM cell** with **non-zero bias**.

Time (t)	Input (x□)
1	1
2	3
3	5

Initial Conditions

$$h_0 = 0, C_0 = 0$$

Given weights and biases:

Gate	Weight (W)	Bias (b)
Forget (f)	0.6	0.1
Input (i)	0.7	0.2
Cell (c)	0.5	0
Output (o)	0.8	0.1

Equations used:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f)$$

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i)$$

$$\tilde{C}_t = \tanh(W_c[h_{t-1}, x_t] + b_c)$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o)$$

$$h_t = o_t \cdot \tanh(C_t)$$

Clustering

- Methods: Distance, Similarity based
- Data Type: Numerical/Categorical
- Numerical Data Type: Hierarchical and Partition based approach

In partition based approach, we have K-Means, K-Medoid (PAM, CLARA, CLARANS).

In Hierarchical approach we have divisive algorithms, DBSCAN, BIRCH, Cure, etc.

Clustering (K-Medoid) cont..

PAM:

i	x	y
X1	2	6
X2	3	4
X3	3	8
X4	4	7
X5	6	2
X6	6	4
X7	7	3
X8	7	4
X9	8	5
X10	7	6

Clustering (DBSCAN) cont..

DBSCAN Numerical 1:
Min Points->4, epsilon->1.9, distance metric-> Euclidian distance

Data Points:

P1: (3, 7) P2: (4, 6)

P3: (5, 5) P4: (6, 4)

P5: (7, 3) P6: (6, 2)

P7: (7, 2) P8: (8, 4)

P9: (3, 3) P10: (2, 6)

P11: (3, 5) P12: (2, 4)

Clustering (DBSCAN) cont..

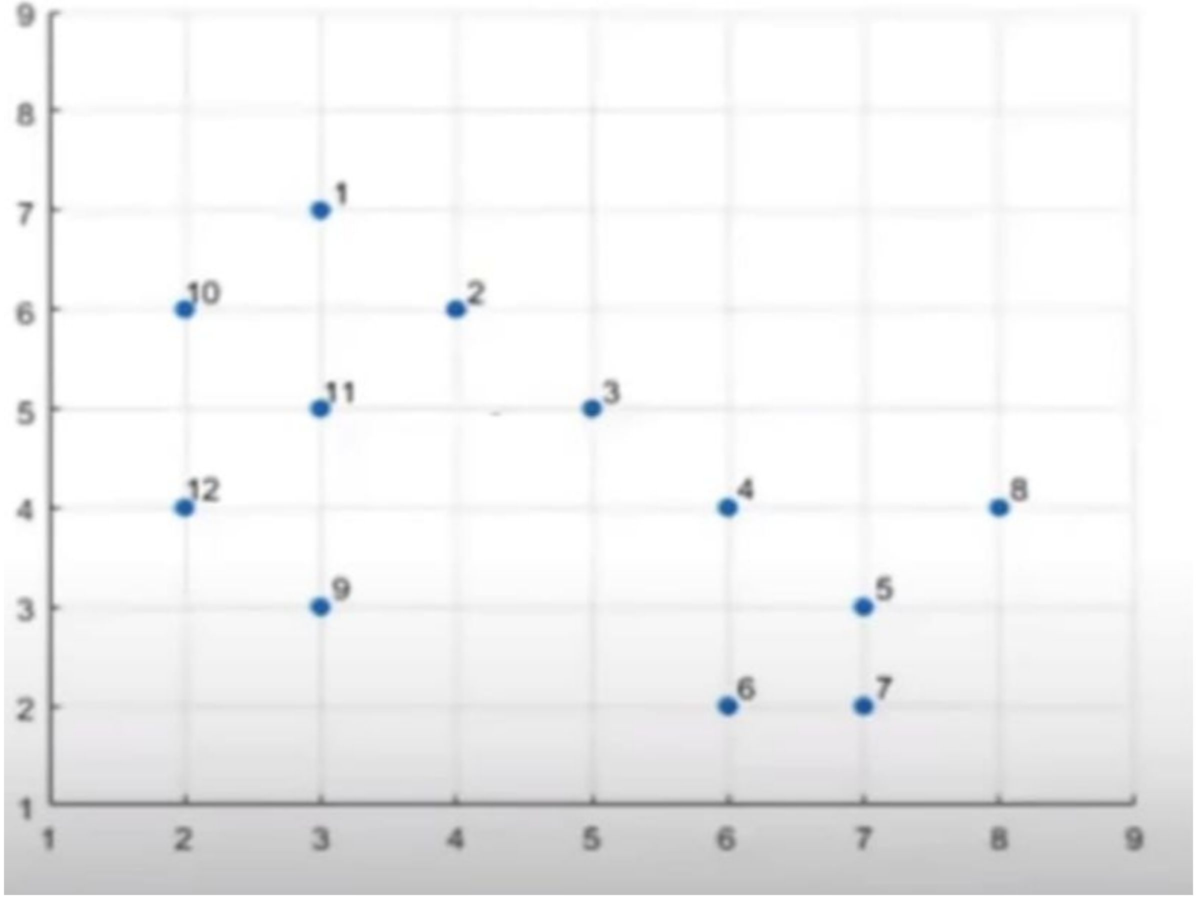
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12
P1	0											
P2												
P3												
P4												
P5												
P6												
P7												
P8												
P9												
P10												
P11												
P12												

Clustering (DBSCAN) cont..

P1: (3, 7)
P2: (4, 6)
P3: (5, 5)
P4: (6, 4)
P5: (7, 3)
P6: (6, 2)
P7: (7, 2)
P8: (8, 4)
P9: (3, 3)
P10: (2, 6)
P11: (3, 5)
P12: (2, 4)

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12
P1	0											
P2	1.41	0										
P3	2.83	1.41	0									
P4	4.24	2.83	1.41	0								
P5	5.66	4.24	2.83	1.41	0							
P6	5.83	4.47	3.16	2.00	1.41	0						
P7	6.40	5.00	3.61	2.24	1.00	1.00	0					
P8	5.83	4.47	3.16	2.00	1.41	2.83	2.24	0				
P9	4.00	3.16	2.83	3.16	4.00	3.16	4.12	5.10	0			
P10	1.41	2.00	3.16	4.47	5.83	5.66	6.40	6.32	3.16	0		
P11	2.00	1.41	2.00	3.16	4.47	4.24	5.00	5.10	2.00	1.41	0	
P12	3.16	2.83	3.16	4.00	5.10	4.47	5.39	6.00	1.41	2.00	1.41	0

Clustering (DBSCAN) cont..



Clustering (DBSCAN) cont..

Numerical 2:

- Min Points: 2
- Similarity Index: 0.8

	P1	P2	P3	P4	P5
P1	1.00	0.10	0.41	0.55	0.35
P2	0.10	1.00	0.64	0.47	0.98
P3	0.41	0.64	1.00	0.44	0.85
P4	0.55	0.47	0.44	1.00	0.76
P5	0.35	0.98	0.85	0.76	1.00

BIRCH Clustering

Balanced Iterative reducing and clustering using hierarchies.

Numerical 2:

$\boldsymbol{x_1} = (\mathbf{3}, \mathbf{4}),\ \boldsymbol{x_2} = (\mathbf{2}, \mathbf{6}),\ \boldsymbol{x_3} = (\mathbf{4}, \mathbf{5}),\ \boldsymbol{x_4} = (\mathbf{4}, \mathbf{7}),\ \boldsymbol{x_5} = (\mathbf{3}, \mathbf{8}),\ \boldsymbol{x_6} = (\mathbf{6}, \mathbf{2}),$

$\boldsymbol{x_7} = (\mathbf{7}, \mathbf{2}),\ \boldsymbol{x_8} = (\mathbf{7}, \mathbf{4}),\ \boldsymbol{x_9} = (\mathbf{8}, \mathbf{4}),\ \boldsymbol{x_{10}} = (\mathbf{7}, \mathbf{9})\ \boldsymbol{T} < \mathbf{1.5}$**, and** $\boldsymbol{Max\ Branch} = \mathbf{2}$

BIRCH Clustering cont..

- Balanced Iterative reducing and clustering using hierarchies.

Dataset:

Point	X	Y
P1	2	3
P2	3	4
P3	4	5
P4	10	12
P5	11	13

For K=3, Apply PAM and do the clustering.

	P1	P2	P3	P4	P5	P6	P7
P1	1.0	0.9	0.4	0.3	0.2	0.1	0.1
P2	0.9	1.0	0.5	0.4	0.3	0.2	0.1
P3	0.4	0.5	1.0	0.8	0.3	0.2	0.2
P4	0.3	0.4	0.8	1.0	0.4	0.3	0.2
P5	0.2	0.3	0.3	0.4	1.0	0.9	0.8
P6	0.1	0.2	0.2	0.3	0.9	1.0	0.9
P7	0.1	0.1	0.2	0.2	0.8	0.9	1.0

	P1	P2	P3	P4	P5	P6	P7	P8	P9
P1	1.0	0.9	0.7	0.2	0.2	0.1	0.1	0.1	0.1
P2	0.9	1.0	0.8	0.3	0.2	0.2	0.1	0.1	0.1
P3	0.7	0.8	1.0	0.4	0.3	0.2	0.2	0.1	0.1
P4	0.2	0.3	0.4	1.0	0.8	0.6	0.2	0.1	0.1
P5	0.2	0.2	0.3	0.8	1.0	0.7	0.2	0.1	0.1
P6	0.1	0.2	0.2	0.6	0.7	1.0	0.3	0.2	0.1
P7	0.1	0.1	0.2	0.2	0.2	0.3	1.0	0.9	0.8
P8	0.1	0.1	0.1	0.1	0.1	0.2	0.9	1.0	0.9
P9	0.1	0.1	0.1	0.1	0.1	0.1	0.8	0.9	1.0

For K=3, Apply PAM and do the clustering.

Point	X	Y
P1	2	3
P2	3	4
P3	4	5
P4	10	12
P5	11	13
P6	12	14

For T=2.5, and Branch=3; apply BIRCH Clustering algorithm and form the clusters

Point	X	Y
P1	2	2
P2	3	3
P3	4	4
P4	8	8
P5	9	9
P6	15	15
P7	5	2

- T=2.5, B=2, Apply BIRCH
- Epsilon=2.5, MinPts=2, distance metric-> Euclidian distance Apply DBSCAN

Analyse the difference.