Machine Learning

SLE Presentation

- SVM,
- Decision Tree,
- Logistic Regression

Presented By:
Group 12

Group Members :-

56

Neha
Tannir

57

Asmi
Mudras

58

Ayush
More

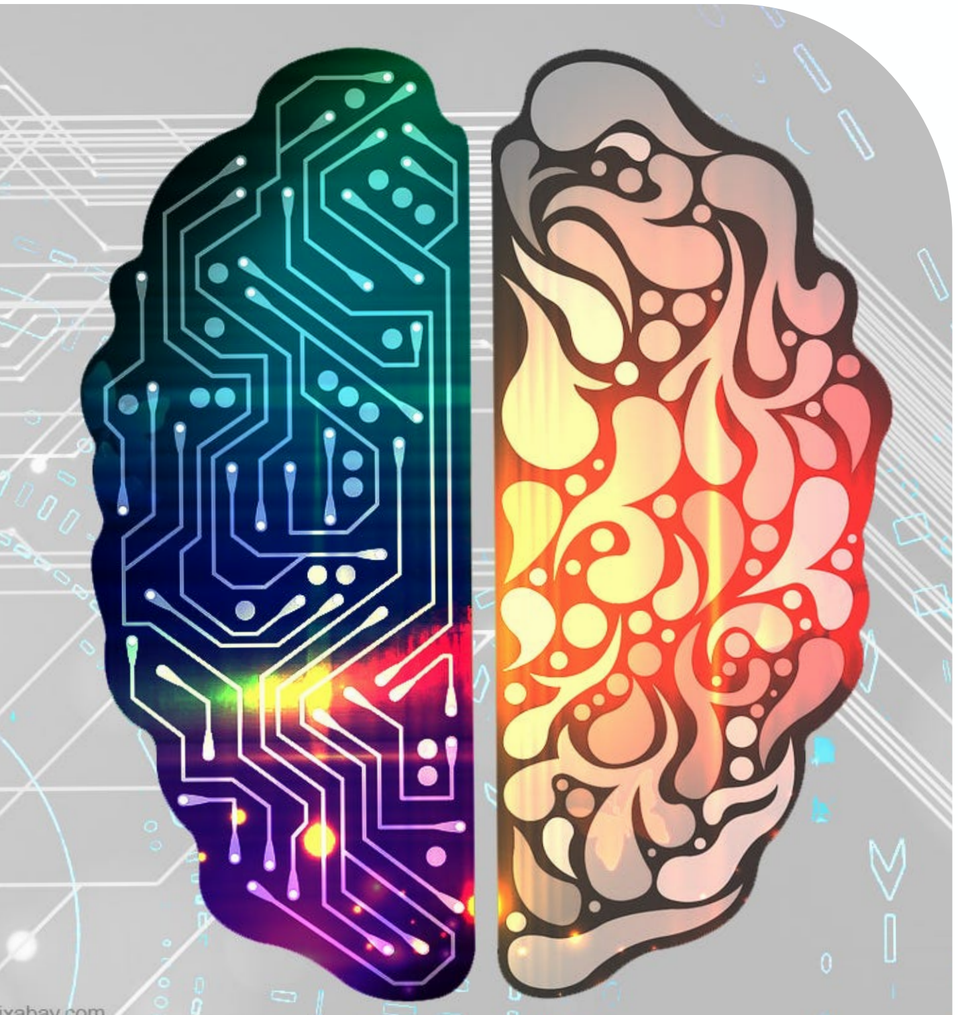
59

Dnyaneshwari
Gund

60

Shadab
Shaikh

What is Machine Leaning

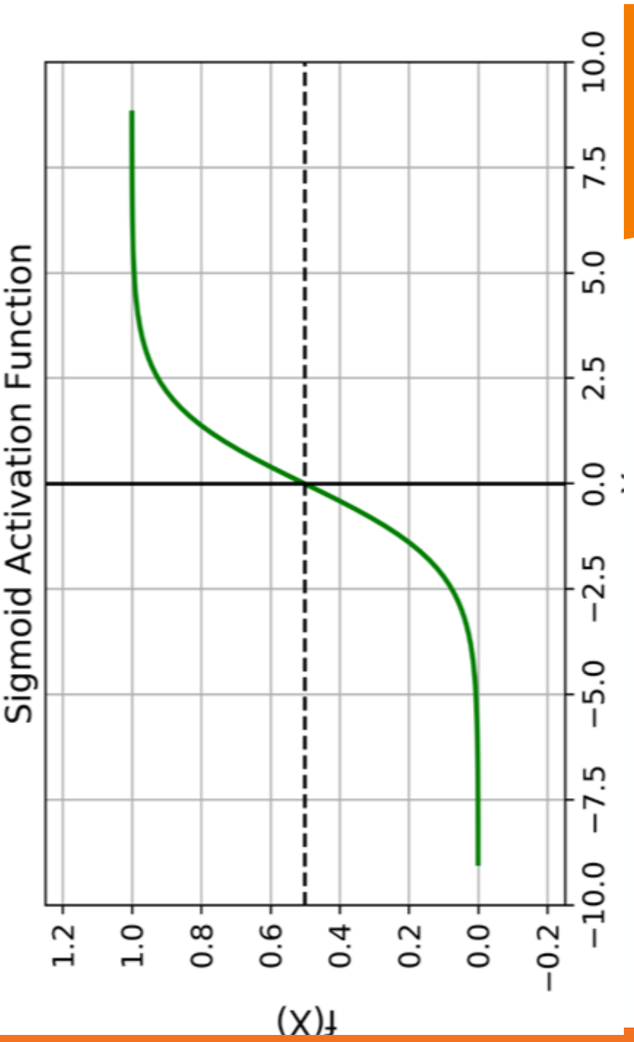


Definition of Machine Learning:

Machine Learning is a branch of artificial intelligence (AI) that enables systems to learn from data and improve over time without being explicitly programmed.

Key Concepts:

- Data: Machine Learning algorithms require large volumes of data to learn patterns and make predictions or decisions.
- Algorithms: These are the mathematical models or techniques used by machines to learn from data.
- Automation: Machine Learning automates the process of discovering insights or making predictions, thereby enhancing efficiency and accuracy.



Logistic Regression

Logistic Regression is a fundamental algorithm in machine learning used for Logistic regression predicts the probability of a binary outcome, like "yes" or "no", by modeling the relationship between dependent and independent variables using a logistic function.

01

Simple and Interpretable: Easy to understand and interpret model coefficients.

02

Computational efficiency for large datasets.

03

Provides probabilities for class predictions.

01

Assumes a linear relationship between variables, which might not always be the case.

02

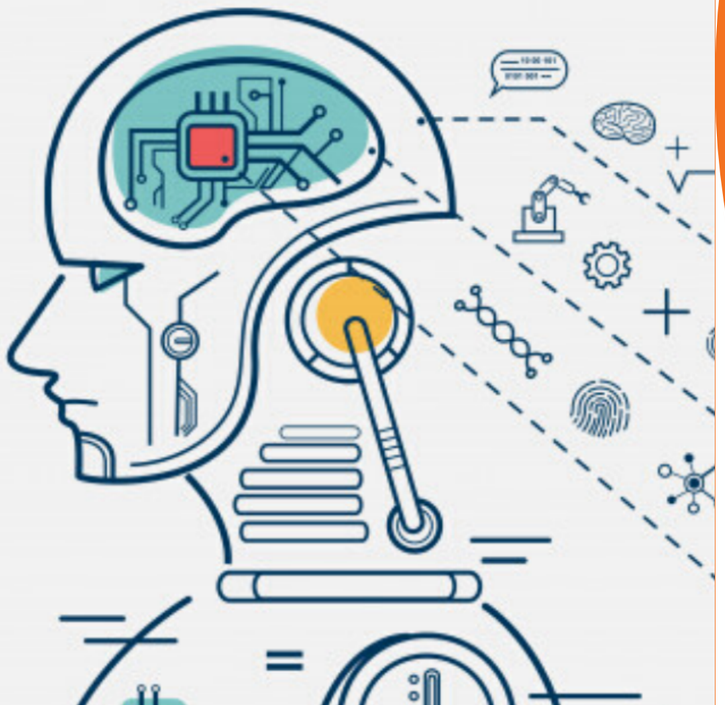
Sensitive to outliers. Outliers can bias predictions.

03

Cannot capture complex relationships between variables.

Working

Linear Regression



Data Preparation:

- Features are preprocessed (e.g., scaled, encoded).
- Data is split into training and testing sets.

Model Training:

- Logistic Regression model is initialized.
- Model parameters (coefficients) are iteratively optimized to maximize the likelihood of predicting the correct class labels.

Prediction

- Given new data, logistic function is applied to calculate probabilities of class membership.
- Decision threshold is applied to assign class labels.

Model Evaluation:

- Model performance is assessed using evaluation metrics (e.g., accuracy, precision, recall).
- Model is fine-tuned if necessary.

Interpretation:

- Coefficients of features indicate their impact on the predicted probabilities.
- Feature importance is assessed to understand the model's decision-making process.

Algorithm (Mathematics Behind This)

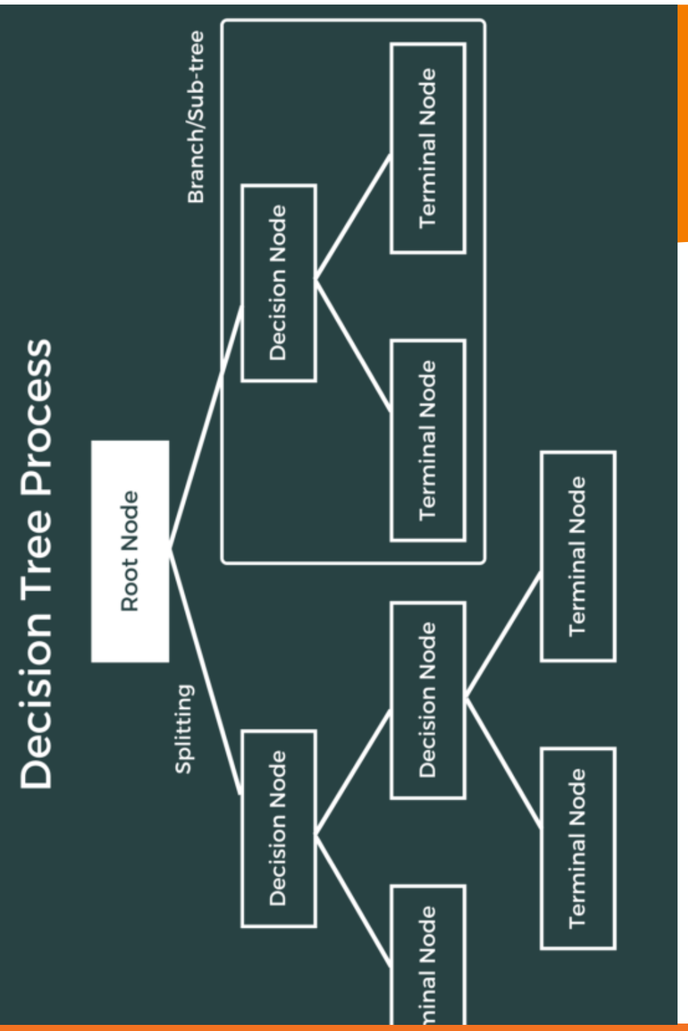
Mathematical Formulation:

The linear regression model can be represented as:

$Y = \beta 0 + \beta 1X1 + \beta 2X2 + \ldots + \beta nXn + \varepsilon$ where,

- Y is the predicted value (target)
- β0 is the intercept
- β1, β2, ..., βn are the coefficients of the features X1, X2, ..., Xn
- ε represents the error term

Decision Tree Classifier



Decision Tree Classifier is a supervised learning algorithm used for classification tasks. A Decision Tree Classifier is a machine learning algorithm that creates a tree-like structure to classify data by making a series of decisions based on feature values. It recursively splits the data into subsets until it reaches a stopping condition, resulting in a set of rules for classification.

01

Interpretability

02

Handling of Both Numerical and Categorical Data

03

Non-parametric

01

Prone to overfitting, especially with complex trees.

02

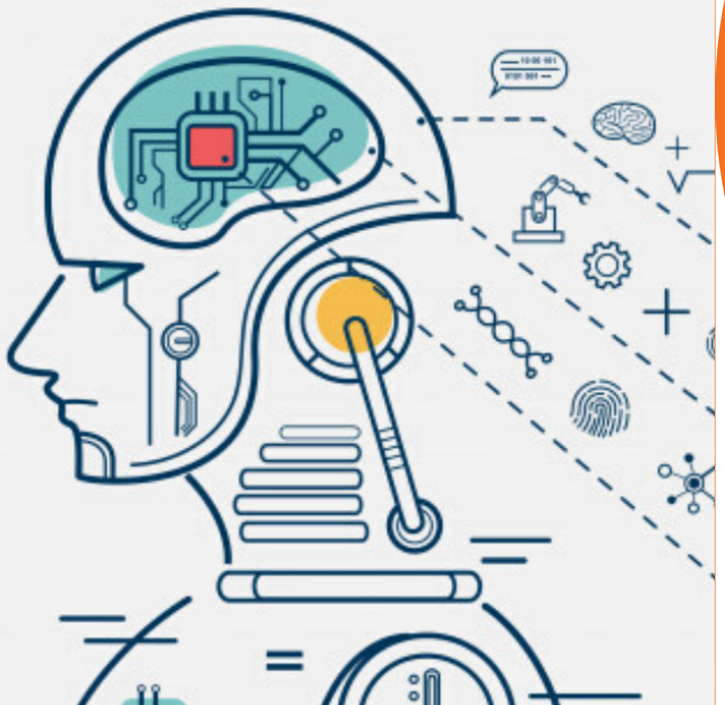
Instability: Small variations in the data can result in a completely different tree.

03

Limited expressiveness: Decision trees may not capture complex relationships as effectively as other algorithms.

Working

Decision Tree Regression



1.Recursion.

2.Minimize Impurity.

3.continues until stopping criterion.

4. Majority Voting

The algorithm recursively selects the best feature to split the dataset based on metrics like information gain or Gini impurity, creating branches of the tree.

At each node, the algorithm aims to minimize impurity by selecting the feature and its corresponding threshold that best separates the data into pure subsets.

This process continues until a stopping criterion is met, such as reaching a maximum tree depth or having a minimum number of samples in a node.

Once the tree is constructed, predictions are made by assigning the majority class of the training samples within each leaf node.

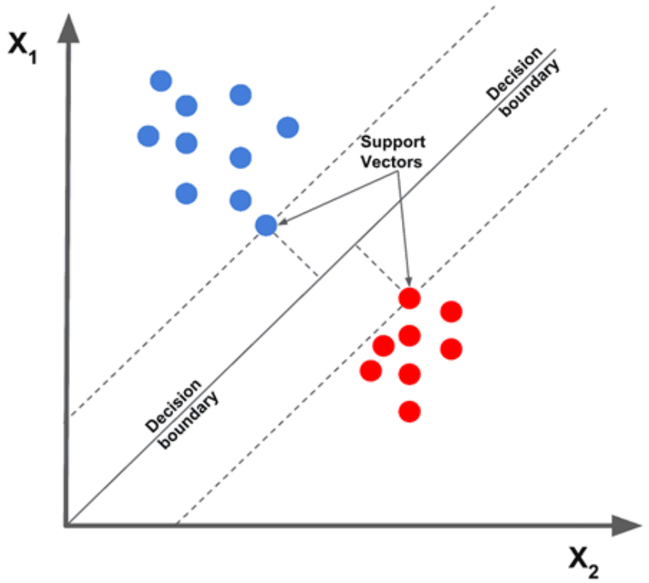
Algorithm

Example :

- Given a dataset with a single feature X and a target variable y.
- The algorithm recursively partitions the feature space based on X, minimizing impurity at each step.
- When predicting the target value for a new data point:
  - a. Traverse the tree based on the value of X.
  - b. Take the majority class label of y in the corresponding leaf node as the predicted class.

Support Vector Machine

•Support Vector Machine (SVM) is a supervised machine learning algorithm used for both classification and regression.
•It is used for text classification, spam detection, image recognition, etc.



01
Effective in high-dimensional spaces

02
Resistant to overfitting & Versatile

03
Effective in cases of limited data

01
Computationally intensive & Sensitive to parameter tuning

02
Lack of probabilistic outputs & Scalability issues

03
Difficulty in interpreting complex models

Algorithm

•It finds the optimal hyperplane in an N-dimensional space that can separate the data points in different classes in the feature space.

• Hyperplane is the decision boundary that is used to separate the data points of different classes in a feature space.

•The dimension of the hyperplane depends upon the number of features. It is a line in 2D and a plane in 3D.

```
from sklearn import svm

# features
X = [
    [-3, -1],
    [0, -2],
    [-2.5, 2],
    [-1, -1],
    [3, .5],
    [.5, 3],
    [-3, -3],
]


# labels
y = [0, 1, 0, 1, 1, 0, 1]


# fit
clf = svm.SVC(kernel='linear').fit(X, y)


# predict
clf.predict([[2, 4]]) -> 0
```

Reference Links

https://www.w3schools.com/python/pandas/default.asp

https://www.youtube.com/watch?v=iIkJrwVUl1c

https://www.javatpoint.com/machine-learning

https://www.kaggle.com/datasets

Thank You!