

Unit -2 ✓

1. Explain feature selection. Compare Filter & wrapper methods

Feature Selection

It is the process of selecting the most relevant & useful features.

In machine learning, a data set may contain a large no. of features. Not all features are useful for predicting the target variable.

Example:

Suppose we want to predict whether a student will pass an exam. The dataset may contain:

- Attendance
- Internal marks
- Study hours
- Previous Marks
- Student name
- Roll no
- Favourite color

Objectives of feature selection

1. Reduce the no. of features in the dataset
2. Remove irrelevant & redundant features.
3. Improve model performance
4. Reduce overfitting.

Types of feature selection methods

1. Filter methods
2. Wrapper methods
3. Embedded methods

Advantages of Feature selection

- Reduces dimensionality
- Improves model efficiency
- Reduces overfitting
- Makes models easier to interpret

2. Describe The Decision Tree algorithm & its Advantages

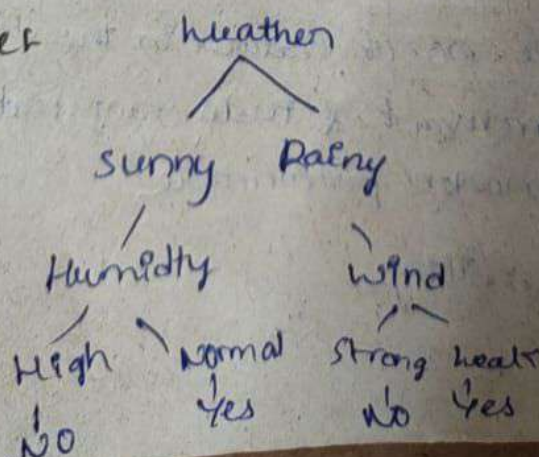
Decision Tree

A Decision tree is a supervised machine learning algorithm used for both classification & regression problems.

The tree consists of

- Root Node
- Branches
- Leaf nodes

Example



Advantages

1. Easy to understand
2. Easy to Interpret
3. Handles classification & Regression
4. Requires Less Data preprocessing
5. Handles Numerical & categorical data

Disadvantages

- They can easily overfit training data
- Very deep trees can become complex
- Large trees may be difficult to interpret.

3. Explain principal component Analysis (PCA) & its Applications.

PCA

It is a dimensionality reduction technique used in machine learning & Statistics to reduce the number of features in a dataset while preserving as much important information as possible.

In a dataset, many features may contain similar or redundant information. PCA transforms the original features into a new set of variables.

Why PCA Needed

- Increase computational cost
- Make visualization difficult
- Increase the possibility of overfitting
- Make the dataset more complex

Advantages:

1. Reduce dimensionality
2. Reduces computational complexity
3. Removes Redundancy
4. Helps redundancy
5. Can reduce noise

Applications of PCA

1. Image processing
2. Data visualization
3. Machine learning
4. Face Recognition

Limitations of PCA

- It is sensitive of feature scaling
- PCA can make the original features harder to interpret
- PCA is mainly designed to capture linear Relationships.