

Practical-1

Machine learning basics:

In this lab, we will go through the basics of machine learning. The student needs to make a soft copy note on the following topics:

Topics:

1. What is Machine learning

Machine learning is a subfield of artificial intelligence (AI) that involves the development of algorithms and models which enable computers to learn from data and make predictions or decisions without explicit programming. It encompasses various techniques and approaches for tasks such as pattern recognition, classification, regression, and clustering.

2. Steps in collection of data

- Define the objectives and goals of data collection.
- Identify the data sources and gather relevant data.
- Clean and preprocess the data to ensure quality.
- Organize and store the data in a suitable format for analysis.
- Ensure data privacy and compliance with regulations.

3. Steps in importing the data in python (Through: csv, json, and other data formats)

- Import necessary libraries (e.g., Pandas for data manipulation).
- Use library-specific functions to read data from CSV, JSON, or other formats.
- Store the data in data structures like DataFrames for further analysis.

➤ Importing Data from CSV:

Step 1: Import the necessary library (e.g., Pandas).

Step 2: Use the Pandas function `read_csv()` to load data from the CSV file into a DataFrame.

```
import pandas as pd
# Step 2: Import data from CSV
csv_file_path = 'your_file.csv'
data = pd.read_csv(csv_file_path)
```

➤ Importing Data from JSON:

Step 1: Import the necessary library (e.g., Pandas).

Step 2: Use the Pandas function `read_json()` to load data from the JSON file into a DataFrame.

```
import pandas as pd
# Step 2: Import data from JSON
json_file_path = 'your_file.json'
data = pd.read_json(json_file_path)
```

4. Preprocessing

a) Remove Outliers:

Identify and remove data points that are significantly different from the majority of the data.

b) Normalize Datasets, Data encoding:

Scale numerical data to a standard range, and encode categorical data into a numerical format.

c) Handling Missing Data:

Fill in or impute missing values using techniques like mean imputation or interpolation.

5. Machine Models

a) Types of machine learning models – Supervised learning, Unsupervised learning, rein

- Supervised Learning: Learn from labeled data for prediction (e.g., regression, classification).
- Unsupervised Learning: Discover patterns and structures in unlabeled data (e.g., clustering, dimensionality reduction).
- Reinforcement Learning: Learn optimal actions through interaction with an environment.

b) Parameters of machine learning model (Learning rate, regularization, etc.)

Parameters include learning rate (controls the step size during training), regularization (to prevent overfitting), and others specific to the model being used.

6. Test-train data split: using constant ration, k-fold cross validation

Using Constant Ratio: Split the dataset into a training set and a testing set with a fixed percentage allocation.

K-Fold Cross Validation: Divide the data into 'k' subsets, use each subset as the test set while the others are used for training, and repeat this process 'k' times to obtain robust performance estimates.

7. Output Inference

Analyze the model's predictions or decisions based on the test data to assess its performance and suitability for the task.

8. Validation: different metrics – Confusion Matrix, Precision, Recall, F1-score

- Confusion Matrix: Provides insights into true positives, true negatives, false positives, and false negatives.
- Precision: Measures the accuracy of positive predictions.
- Recall: Measures the ability to identify all relevant instances.
- F1-Score: Combines precision and recall to assess model performance.