Training Day 21 Report

Date: 19 July 2025
Topic: Linear Regression with Pipeline and Model Saving

Overview

Today's session focused on implementing **Linear Regression** using the **scikit-learn Pipeline** and saving trained models for future use. Pipelines help streamline workflows by chaining multiple steps (like preprocessing and modeling) together, while model saving ensures reusability without retraining.

Key Concepts

1. Linear Regression

- A supervised learning algorithm used for predicting continuous values.

- Finds the best-fit line using the formula:
  $y = mX + c$

2. Pipeline in Scikit-Learn

- A tool to combine multiple steps into one object.

- Example steps:

  - **StandardScaler()** – scales features for better performance.

  - **LinearRegression()** – fits regression model on scaled data.

3. Model Training and Workflow

- Dataset generated with make_regression() function.

- Pipeline fitted with scaling + regression.

- Model learns relationship between input features and output.

4. Model Saving with Joblib

- Saves trained model into a .pkl file.

- Allows reuse of the model without retraining.

Example:

```
joblib.dump(pipeline, "linear_regression.pkl")
model = joblib.load("linear_regression.pkl")
```

Visualization: Pipeline Workflow

Input Data → [StandardScaler] → [Linear Regression] → Prediction

Learning Outcome

- Understood the use of **Pipeline** for combining preprocessing and modeling.

- Learned to apply **Linear Regression** on synthetic data.

- Gained knowledge on **saving and loading models** for deployment.