Practice Problems

Tasks:

1. Load and preprocess the dataset (handle categorical variables like Fuel_Type, Seller_Type, Transmission).
2. Perform train-test split (80/20).
3. Train a Linear Regression model.
4. Report RMSE, MAE, and R^2 score.
5. Train Ridge and Lasso regression models with different values of α (alpha).
6. Plot RMSE vs. α to visualize regularization effects.

1. Car Price Prediction

Problem

Predict the price of a car based on attributes like mileage, engine size, horsepower, brand, etc.

Dataset

- Dataset: https://www.kaggle.com/datasets/nehalbirla/vehicle-dataset-from-cardekho
- Target: Selling_Price
- Features: Year, Present_Price, Kms_Driven, Fuel_Type, Seller_Type, Transmission, etc.

Notes

- Handle categorical features with encoding.
- Linear, Ridge, or XGBoost regression work well.

2. Student Performance Prediction

Problem

Predict final exam scores based on study time, parental education, and attendance.

Dataset

- Dataset: Student Performance Data
  - § https://www.kaggle.com/spscientist/students-performance-in-exams
- Target: math score, reading score, or writing score
- Features: gender, lunch, test preparation course, study time, etc.

Notes

- Can practice with Polynomial Features.
- Good for comparing linear vs tree models.

3. Used Laptop Price Prediction

Problem

Predict laptop resale prices based on hardware specs.

Dataset

- Dataset: https://www.kaggle.com/datasets/muhammetvarl/laptop-price
- Target: Price
- Features: RAM, Processor, Brand, Screen Size, Operating System

Notes

- Requires preprocessing of text features (e.g., parsing “Intel i7”).

4. Weather-Based Energy Consumption

Problem

Predict electricity usage based on weather conditions.

Dataset

- Dataset: Energy Efficiency Dataset (UCI)
- Target: Heating Load or Cooling Load
- Features: Relative Compactness, Surface Area, Wall Area, Glazing Area, Orientation, etc.

Notes

- Use feature scaling.
- Try polynomial regression.

📦 5. Bike Sharing Demand

▨ Problem

Predict the number of bikes rented on a given day/hour.

📂 Dataset

- Dataset: Bike Sharing Dataset
- Target: count
- Features: season, holiday, workingday, weather, temp, humidity, windspeed

🔧 Notes

- Great for time series + regression.
- Can use feature engineering on datetime.