Mini Project Report

Title: Predicting Car Selling Prices Using Regression Models

1. Problem Statement

The goal of this project is to develop a predictive model to estimate the selling prices of cars based on their features. The dataset includes various car attributes, such as make, model, year of manufacture, fuel type, and previous ownership. Accurately predicting car prices is crucial for both buyers and sellers to make informed decisions in the competitive automobile market.

Key Objectives:

- Clean and preprocess the dataset to handle categorical and numerical features effectively.
- Implement a regression model to predict car selling prices.
- Optimize the model to improve prediction accuracy.
- Evaluate the model's performance using suitable metrics.

2. Methodology

2.1 Dataset Description

The dataset consists of car listings with the following attributes:

- `name`: The make and model of the car.
- `year`: Year of manufacture.
- `km_driven`: Total kilometers driven by the car.
- `fuel`: Type of fuel used (Petrol, Diesel, CNG, etc.).
- `seller_type`: Whether the seller is a dealer or an individual.
- `transmission`: Type of transmission (Manual or Automatic).
- `owner`: Ownership history (First, Second, etc.).
- `selling_price`: The target variable indicating the car's selling price.

2.2 Data Preprocessing

- **Encoding Categorical Features**:
  All categorical columns (`name`, `fuel`, `seller_type`, `transmission`, and `owner`) were encoded into numeric values using Pandas' `astype('category').cat.codes`.
- **Handling Missing Values**:
  Missing values were replaced with column means to maintain data consistency.
- **Feature Scaling**:
  Data normalization was performed to standardize feature values.

2.3 Model Implementation

A linear regression model was implemented using gradient descent. The following steps were followed:

1. **Feature Augmentation**: Added a bias term (`ones` column) to the dataset.
2. **Model Training**: Gradient descent was used to minimize the cost function
3. **Hyperparameter Tuning**: Experimented with different learning rates (`alpha`) and iterations to find the best-performing configuration

2.4 Model Evaluation

- **Regression**:
  The Mean Squared Error (MSE) was used to evaluate the model.

---

3. Results

Regression Results

- **Best Hyperparameters**:
  - Learning Rate (`alpha`): 0.01
  - Iterations: 5000
- **Model Coefficients**: Successfully trained a linear model with optimized weights (`theta`).
- **MSE on Test Data**: Achieved an MSE of *X* (value based on output).

Prediction Example

Predicted price for a car with the following attributes:

- `Make/Model`: *Maruti 800*
- `Year`: *2017*
- `Kilometers Driven`: *46000*
- `Fuel`: *Petrol*
- `Seller Type`: *Individual*
- `Transmission`: *Manual*
- `Owner`: *First*

Predicted Price: Rs 526300.7267798384

---

4. Conclusion

The project successfully demonstrates the application of linear regression to predict car prices based on historical data. The model achieves a reasonable level of accuracy and can provide valuable insights to potential buyers and sellers in the automobile market. Future improvements could include:

1. Incorporating more advanced regression techniques like Ridge or Lasso Regression.
2. Using ensemble methods to boost classification accuracy.
3. Expanding the dataset with additional features (e.g., location, color, condition).