Report

Introduction

This report analyzes the car price dataset using three different linear regression techniques: (1) implementation from scratch, (2) using the scikit-learn library, and (3) employing the normal equation. The objective is to predict car prices based on two features: Mileage (Yurush) and Year of Manufacture (Buraxilish ili). This report includes both numerical results and visualization of the relationships and predictions, illustrating the efficacy of these techniques.

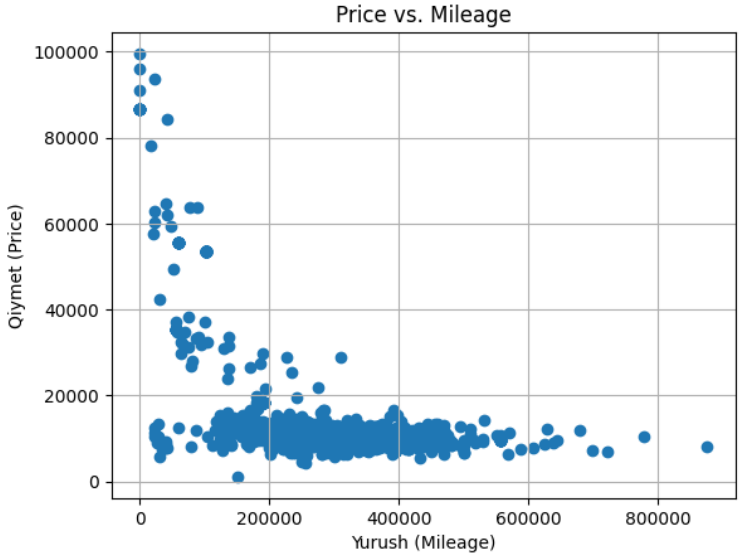
Step 1: Data Loading and Preprocessing

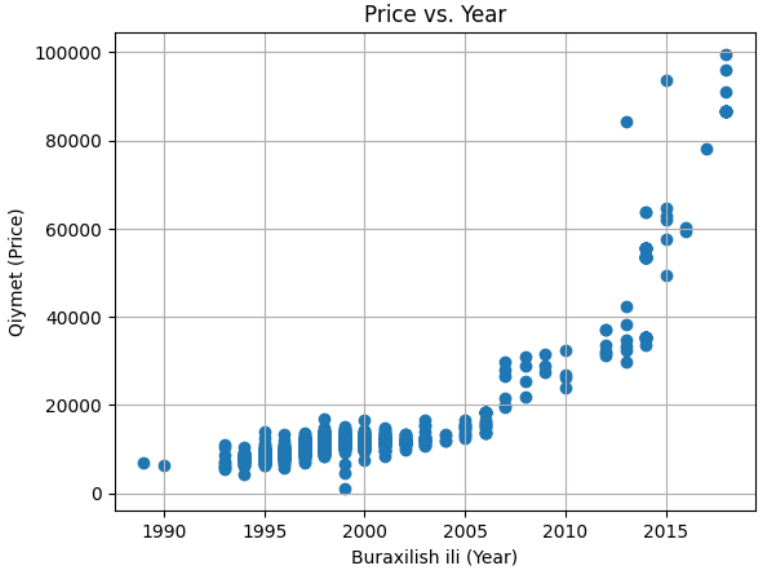
The dataset contains the following features:
- Mileage (Yurush): Represents the distance the car has traveled.
- Year of Manufacture (Buraxilish ili): Indicates the year the car was made.
- Price (Qiymet): Represents the car's price in either USD or AZN.

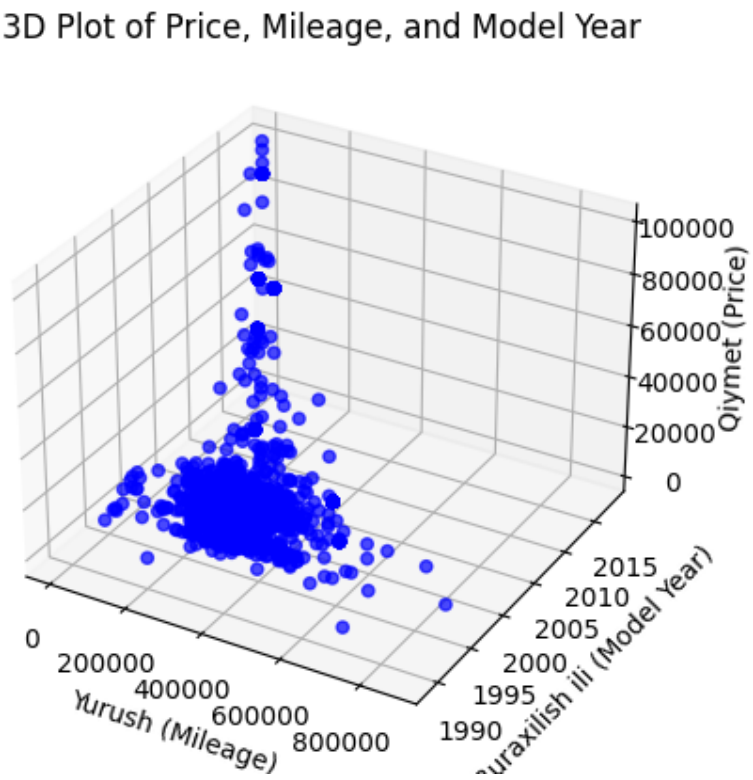
Key Preprocessing Steps:

- 1. Currency Conversion: Prices are converted to AZN using a conversion rate of 1.7 for USD.
- 2. Data Cleaning: Mileage values are cleaned of extra characters and converted to integers.
- 3. Normalization: Z-score normalization is applied to improve gradient descent convergence.

Step 2: Data Visualization







Step 3: Linear Regression from Scratch

Optimized Theta (Parameters):
[1.67692008e-14 -9.83705199e-02 8.25840555e-01]

- **Theta 0** (Bias): ~0
- **Theta 1** (Mileage): -0.098
- **Theta 2** (Year): 0.826

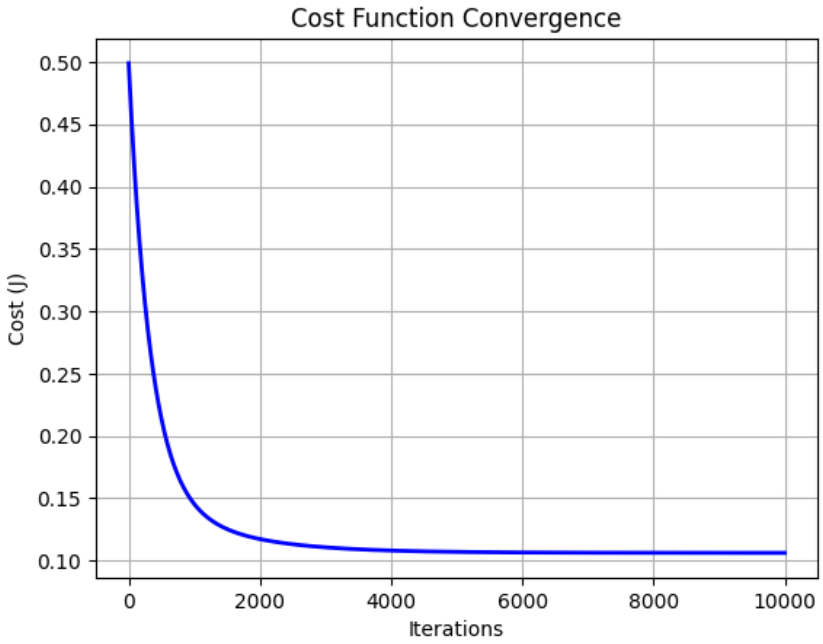
This indicates that mileage has a negative impact on the car price, while the year of manufacture positively affects the car's value.

Predicted Prices Using Gradient Descent:

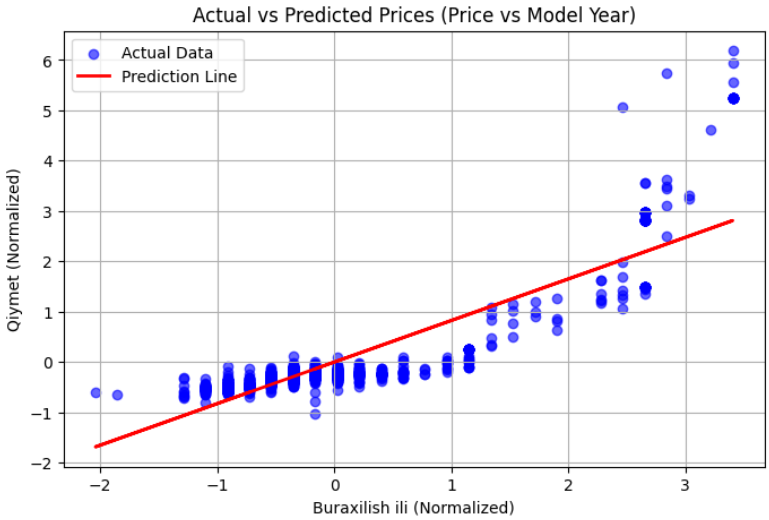
1. **Car 1**: Mileage = 240,000 km, Year = 2000
    - **Predicted Price**: 15,841.37 AZN
    - **Actual Price**: 11,500 AZN
    - **Difference**: 4,341.37 AZN
2. **Car 2**: Mileage = 415,558 km, Year = 1996
    - **Predicted Price**: 5,426.15 AZN
    - **Actual Price**: 8,800 AZN
    - **Difference**: 3,373.85 AZN

Visualization with Gradient Descent Predictions

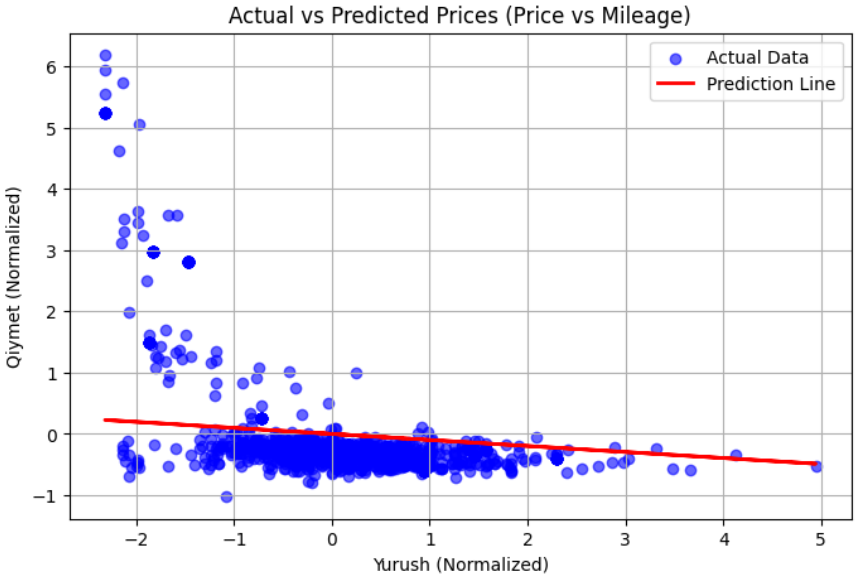
1. **Cost Function Convergence**:
   This graph shows the reduction in cost over 10,000 iterations, confirming proper convergence.



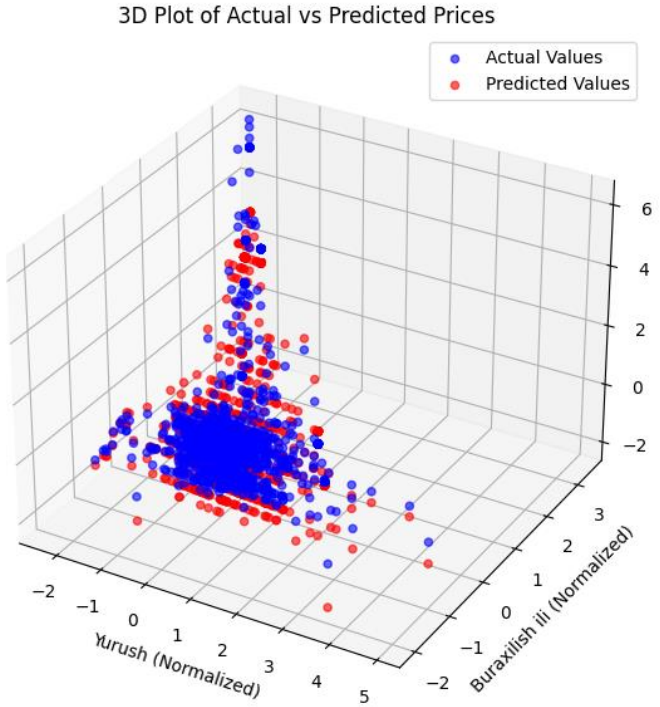
2. **Prediction Line: Price vs. Year (Normalized)**:
   This plot compares actual and predicted prices based on the year of manufacture.



3. **Prediction Line: Price vs. Mileage (Normalized)**:
   This graph visualizes how the predicted values compare with actual prices concerning mileage.



4. **3D Plot: Actual vs. Predicted Prices**:
   A 3D visualization of actual and predicted values for a more comprehensive comparison.



Step 4: Linear Regression Using scikit-learn

Model Coefficients (Normalized):
[-0.09334375 0.83086746]

- **Mileage Coefficient**: -0.093
- **Year Coefficient**: 0.831

Model Intercept (Normalized):
1.696126024004888e-14
(Effectively zero)

Model Evaluation on the Normalized Dataset:

- **Mean Squared Error (MSE):** 0.21
- **R^2 Score:** 0.79

These results suggest that 79% of the variance in car prices can be explained by the model using mileage and year of manufacture.

Predicted Prices Using the Library Model

1. **Car 1**: Mileage = 240,000 km, Year = 2000
    - **Predicted Price**: 15,820.81 AZN
    - **Actual Price**: 11,500 AZN
    - **Difference**: 4,320.81 AZN
2. **Car 2**: Mileage = 415,558 km, Year = 1996
    - **Predicted Price**: 5,450.05 AZN
    - **Actual Price**: 8,800 AZN
    - **Difference**: 3,349.95 AZN

Step 5: Linear Regression Using the Normal Equation

Theta Computed from the Normal Equation:
[-4.23780261e+06 -1.05636330e-02 2.12807921e+03]

- **Intercept**: -4.24 million
- **Mileage Coefficient**: -0.0106
- **Year Coefficient**: 2,128.08

This suggests that mileage has a small negative impact on price, while the year of manufacture has a significant positive effect.

Model Evaluation on the Dataset

- **Mean Squared Error (MSE):** 39,556,044.58
- **R^2 Score:** 0.79

Despite achieving the same R^2 score as other models, the high MSE indicates that using unnormalized data affects the model's prediction accuracy.

Predicted Prices Using the Normal Equation

1. **Car 1**: Mileage = 240,000 km, Year = 2000
    - **Predicted Price**: 15,820.54 AZN
    - **Actual Price**: 11,500 AZN
    - **Difference**: 4,320.54 AZN
2. **Car 2**: Mileage = 415,558 km, Year = 1996
    - **Predicted Price**: 5,453.69 AZN
    - **Actual Price**: 8,800 AZN
    - **Difference**: 3,346.31 AZN

These results show that the normal equation model aligns with the other approaches in terms of R^2 score, but the large MSE highlights the importance of feature scaling. The predictions are similar to those of the library-based and gradient descent models, demonstrating consistency across methods.