Question 1: Linear Regression & Gradient Descent

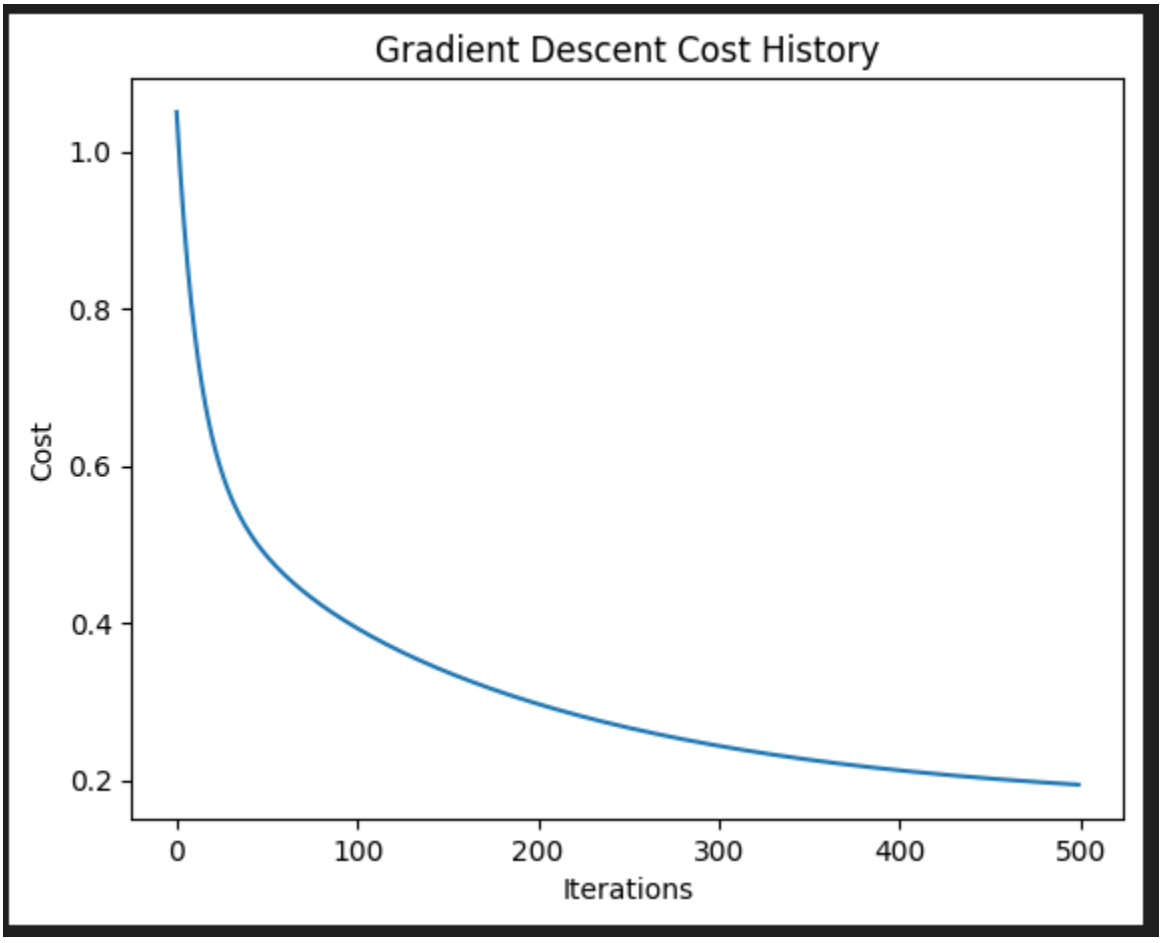
(a) Gradient Descent Implementation

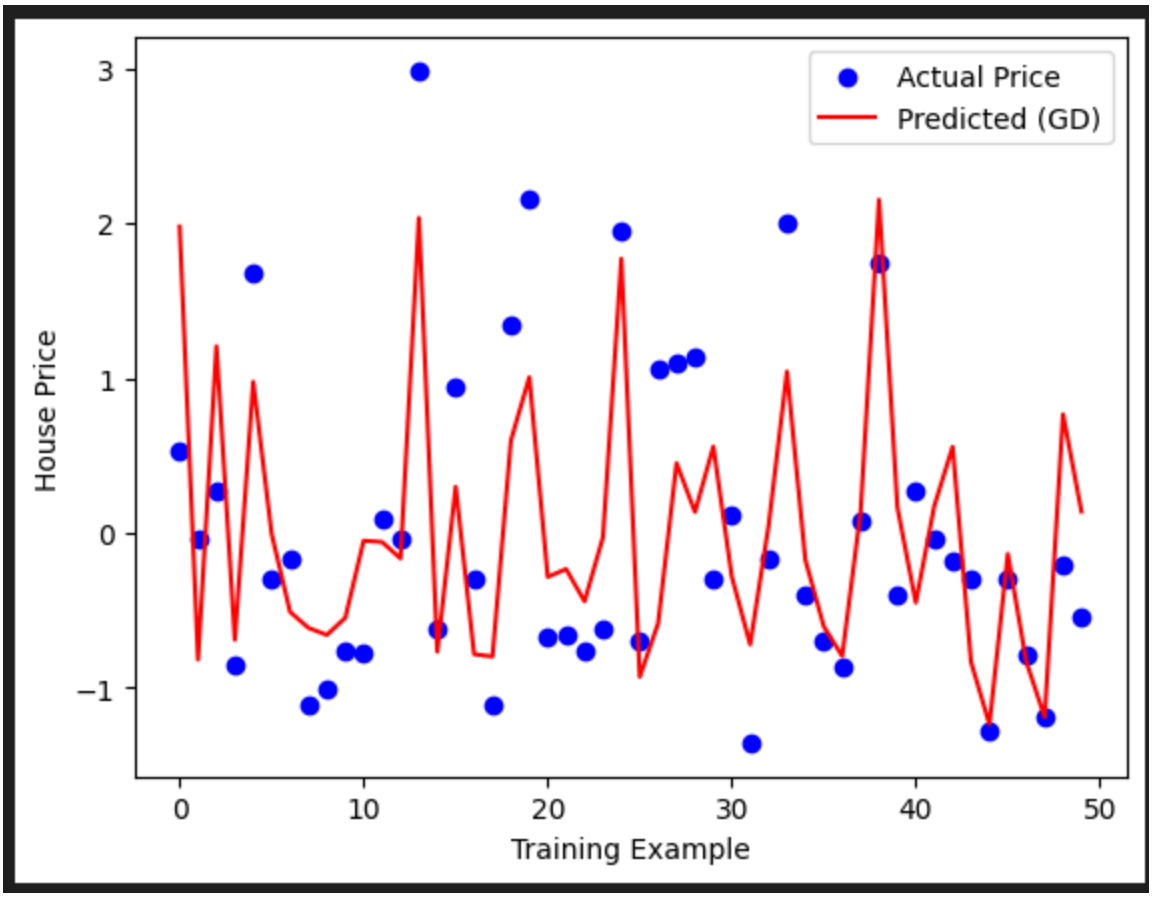
I implemented the gradient descent algorithm with a learning rate of **0.001** after normalizing the data to prevent overflow. The algorithm iteratively updated the parameters until convergence.

- **Final Parameters (θ):**

  θ=[−2.95×10−5, 0.7695, 0.3183, −0.2652]\theta = [-2.95 \times 10^{-5},\ 0.7695,\ 0.3183,\ -0.2652]θ=[−2.95×10−5, 0.7695, 0.3183, −0.2652]

- **Error (Mean Squared Error):**
  0.3409





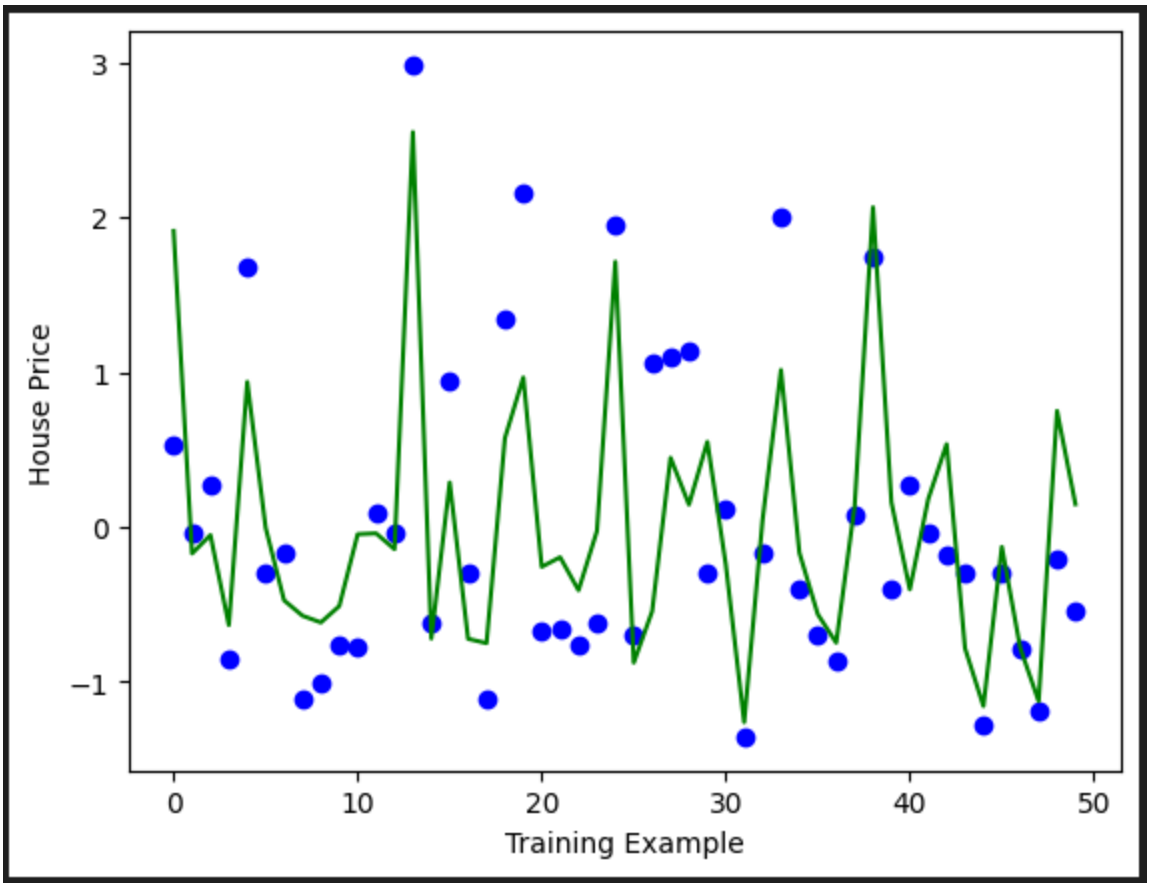
(b) Closed-Form Solution

The closed-form solution was implemented using the normal equation.

- **Final Parameters (θ):**

  θ=[−8.15×10−17, 0.7820, 0.1784, −0.1323]\theta = [-8.15 \times 10^{-17},\ 0.7820,\ 0.1784,\ -0.1323]θ=[−8.15×10−17, 0.7820, 0.1784, −0.1323]

- **Error (Mean Squared Error):**
  0.3360



(c) Comparison of Errors

- Error (Gradient Descent): **0.3409**
- Error (Closed Form): **0.3360**

Both methods produced very similar results. The closed-form had a slightly lower error because it solves the optimization exactly, while gradient descent is iterative and approximate.

(d) Discussion & Conclusion

- **Closed-Form:** Better for small datasets since it directly gives the optimal solution.
- **Gradient Descent:** Useful for large datasets or when the number of features is very high, since closed-form requires matrix inversion (computationally expensive).
- **Conclusion:** In this dataset, both methods worked well, but closed-form performed slightly better in terms of error.

Question 2: Logistic Regression

The logistic regression algorithm was implemented to classify houses into two categories:

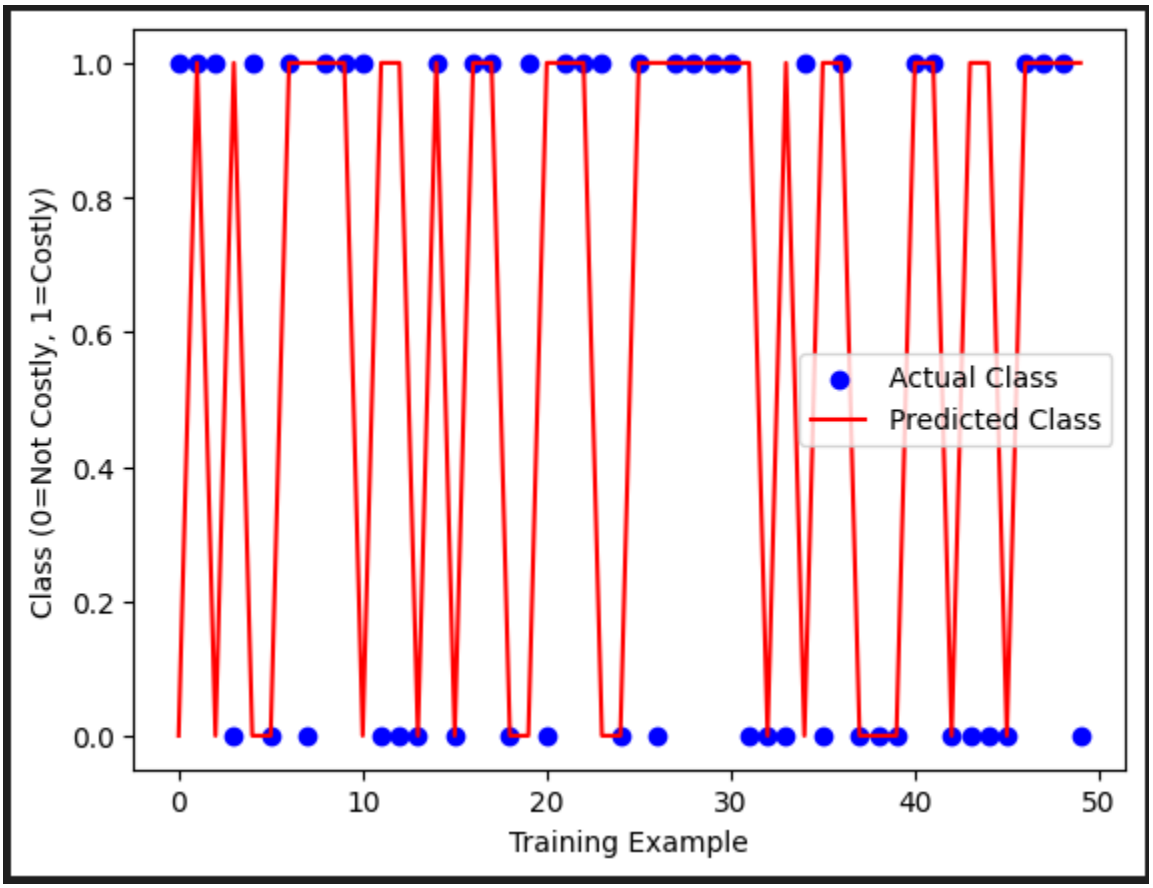
- **1 → Costly**
- **0 → Not Costly**

Results:

- **Final Parameters (θ):**

  θ=[−0.4689, 0.0675, 0.2492, −0.8459]\theta = [-0.4689,\ 0.0675,\ 0.2492,\ -0.8459]θ=[−0.4689, 0.0675, 0.2492, −0.8459]

- **Accuracy: 50%**



Discussion

- Logistic regression trained with gradient descent but achieved **only 50% accuracy**, which means it did not learn meaningful decision boundaries from the given dataset.
- Possible reasons:
  1. Features may not be strong indicators of "Costly" vs "Not Costly".

2. Learning rate or iterations may need fine-tuning.
3. Data may not be linearly separable.

Conclusion

- Linear regression (both GD and closed-form) produced consistent and accurate results for predicting house prices.
- Logistic regression did not perform well, suggesting that either **feature engineering** or a more complex model (like polynomial features or another classifier) may be needed.