Experiments with Learning Rates in Gradient Descent

This report presents the results of testing different learning rates in a Batch Gradient Descent model for linear regression. The experiments explore the effects of using fixed and dynamic learning rates on the model's performance, measured by Mean Squared Error (MSE).

---

1. Objective

The objective of this experiment is to evaluate the impact of different learning rates on the convergence and accuracy of Batch Gradient Descent in linear regression. Specifically, it examines:

- The behavior of the training loss (MSE) with different fixed learning rates.
- The effectiveness of a dynamic learning rate strategy in achieving faster and smoother convergence.

---

2. Experimental Setup

- **Model:** Linear Regression using Batch Gradient Descent.
- **Loss Function:** Mean Squared Error (MSE).
- **Learning Rates Tested:**
  - Fixed: $\alpha = 0.001, 0.01, 0.1$
  - Dynamic: $\alpha = 0.02$, $\alpha = 0.01$ with decay = 0.001, decay step = 10
- **Iterations:** 100

The dynamic learning rate is updated as:

$\alpha = \alpha \ / (1 + \text{decay} \times \text{iterations})$

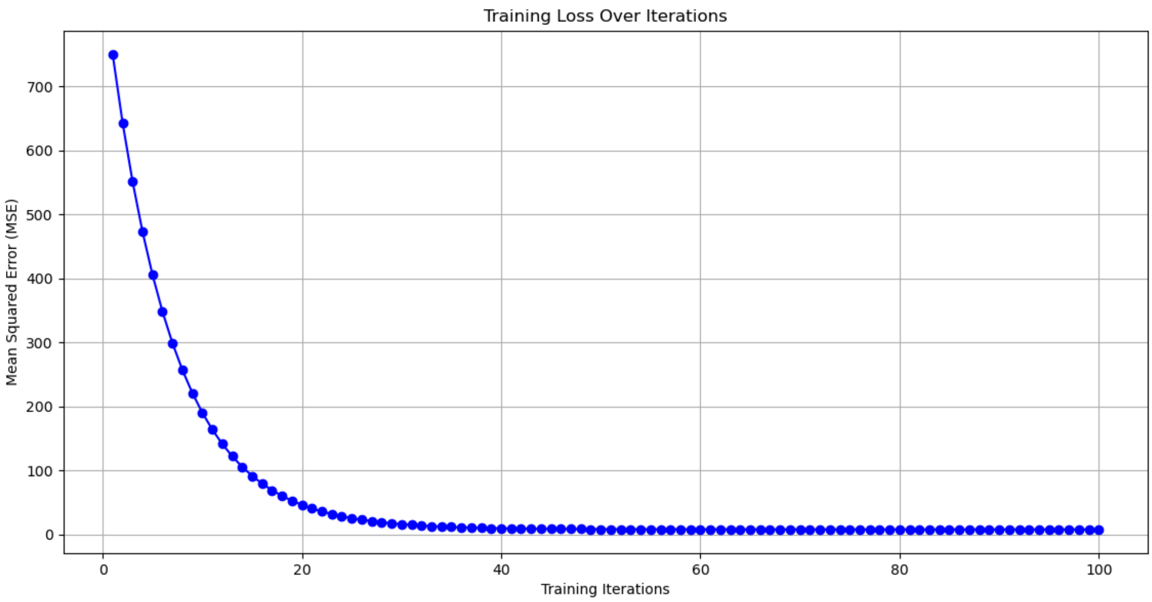
---

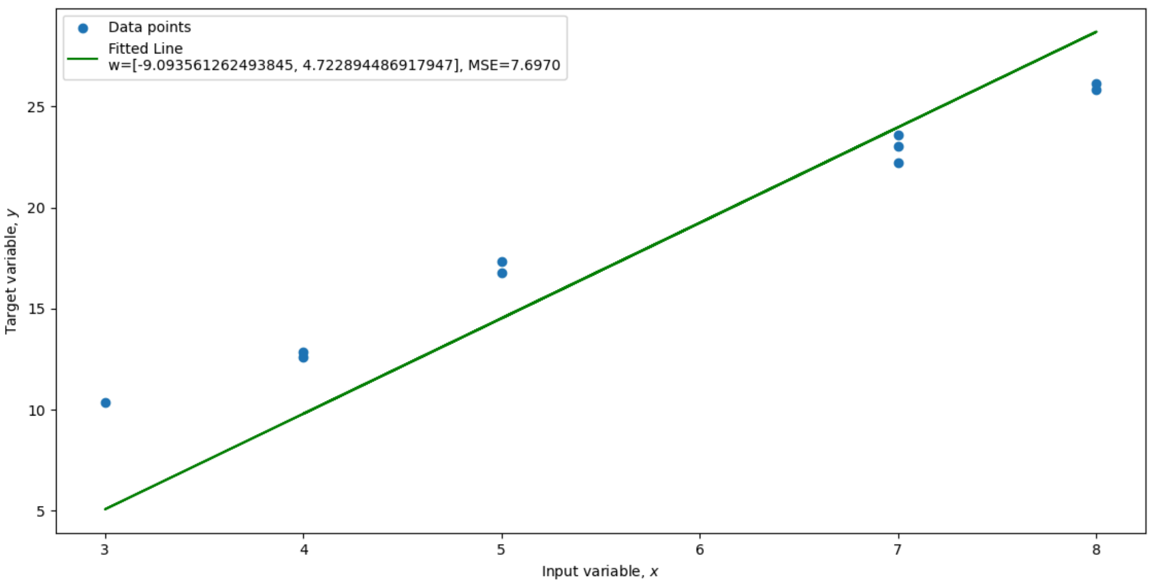
3. Results with Fixed Learning Rates

a. Learning Rate: $\alpha = 0.001$

- **Behavior:** The gradient descent curve is smooth and steadily decreases.
- **Performance:**
  - MSE starts high at 750.188.

- Gradually decreases to mse of 10 after 38 iterations.
- The regression line fits the data well.

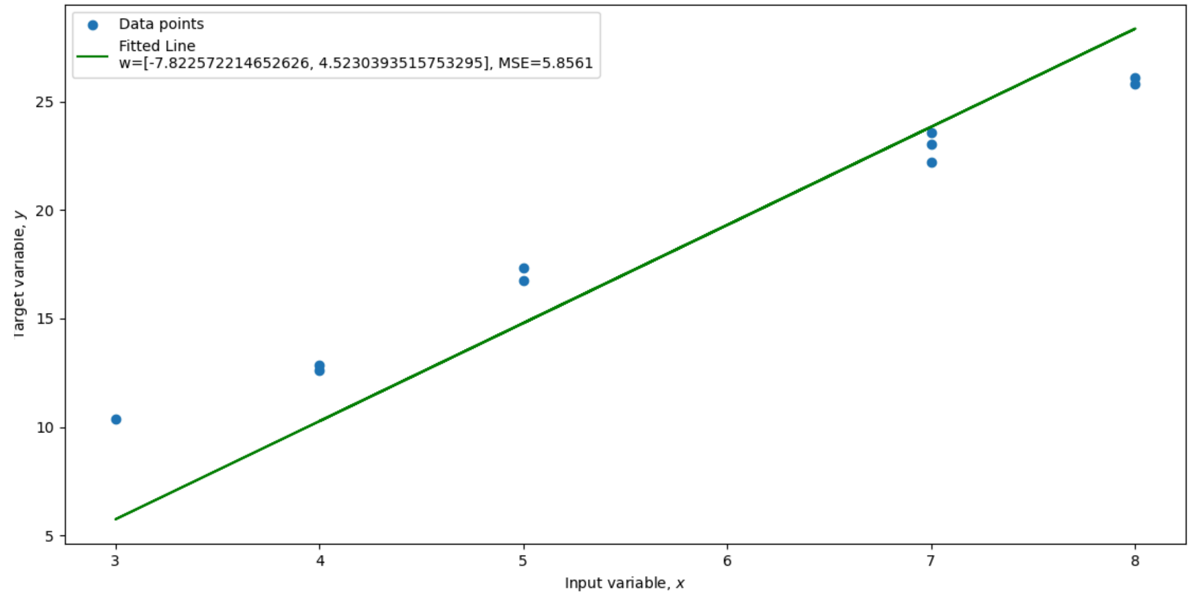


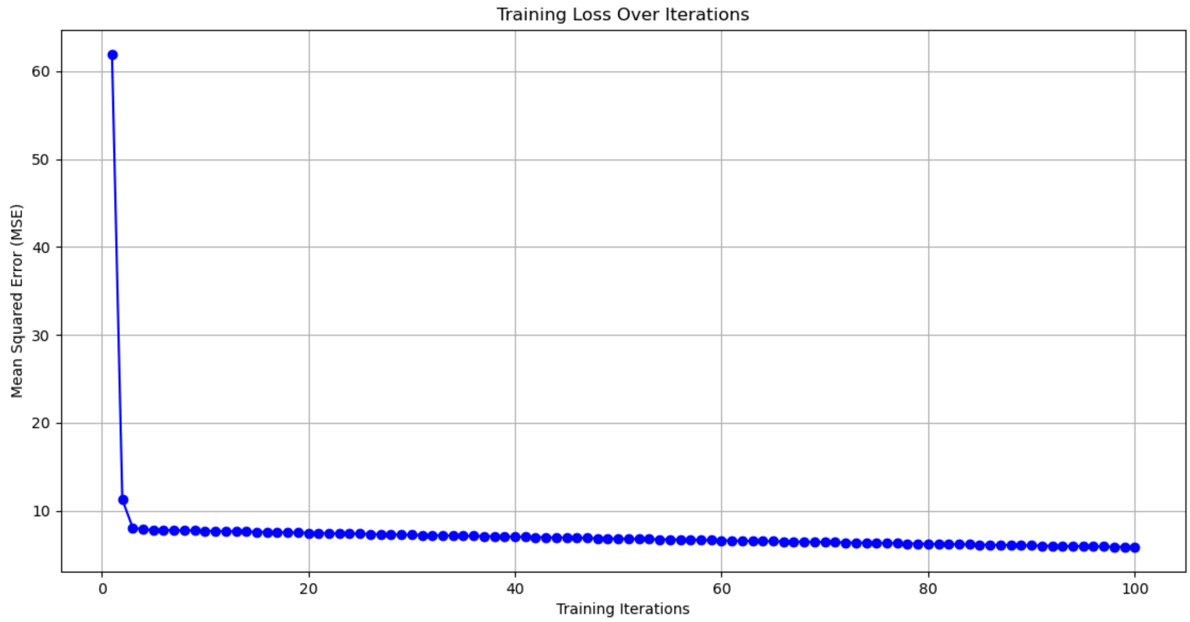


b. Learning Rate: α = 0.01

- **Behavior:** Convergence is faster compared to α = 0.001.
- **Performance:**
  - MSE starts at 62 (much lower than α = 0.001).
  - Sharply decreases to approximately 7 in three iterations.
  - Ends at the lowest MSE of 5.8561.

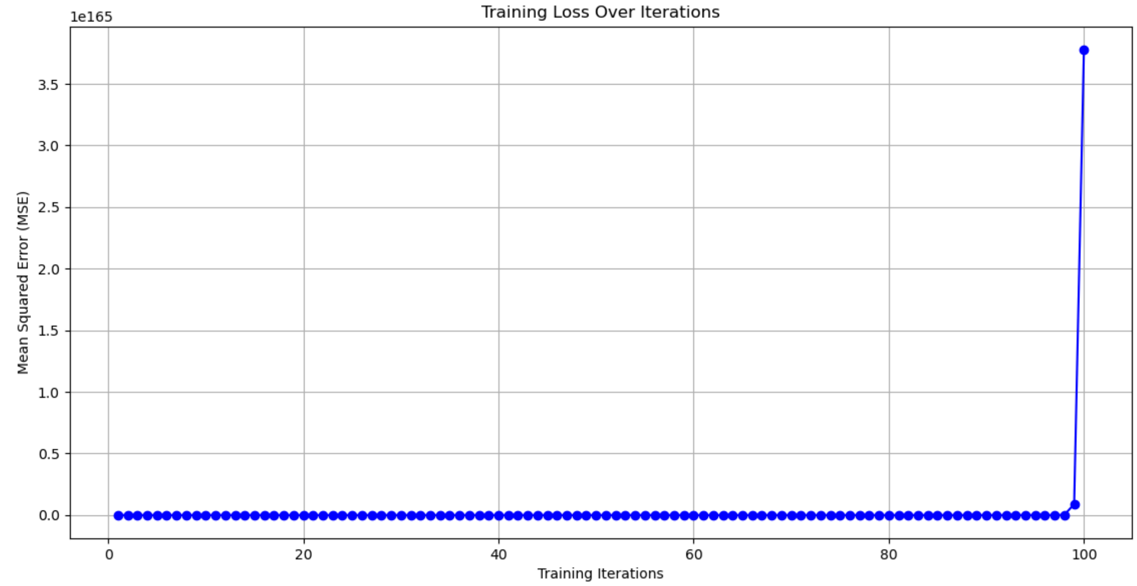
- The line fits the data well with rapid convergence.

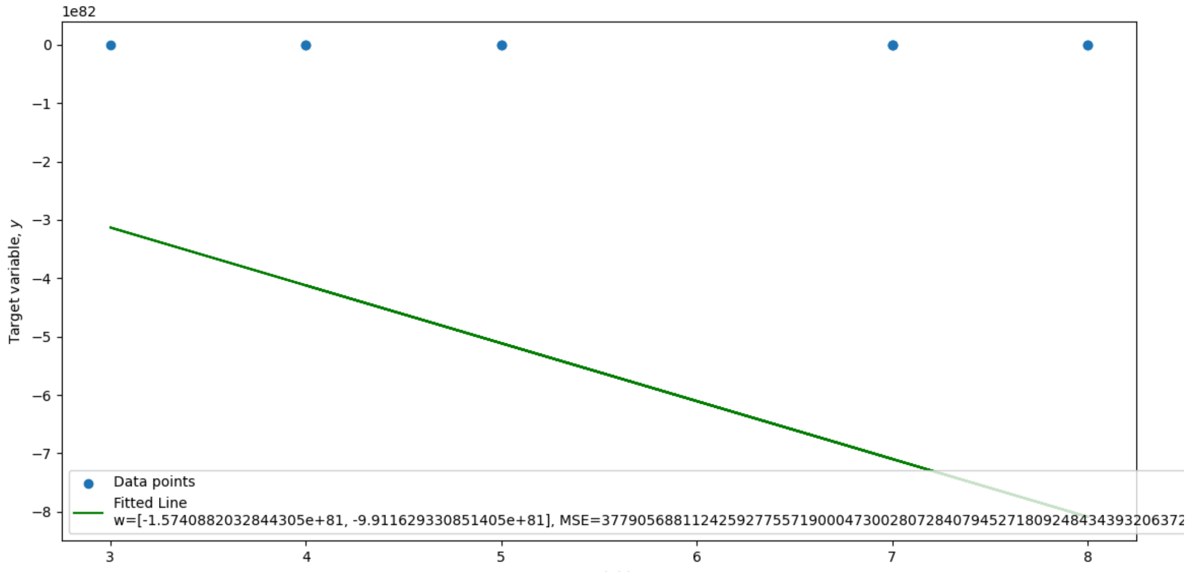




c. Learning Rate: α = 0.1

- **Behavior:** The model diverges and fails to converge.
- **Performance:**
  - MSE starts at 36710.5698 and increases at each iteration.
  - The regression line fails to fit the data.
  - The training loss graph is unusual because of rapidly increasing MSE, indicating the learning rate is too high for stable gradient descent.

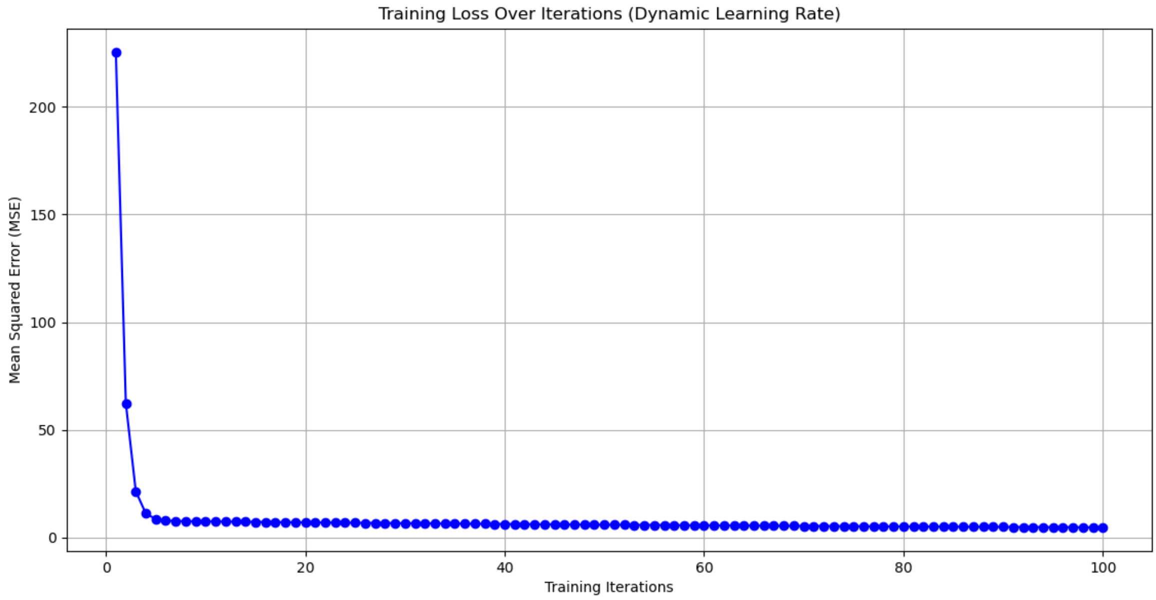




4. Results with Dynamic Learning Rates

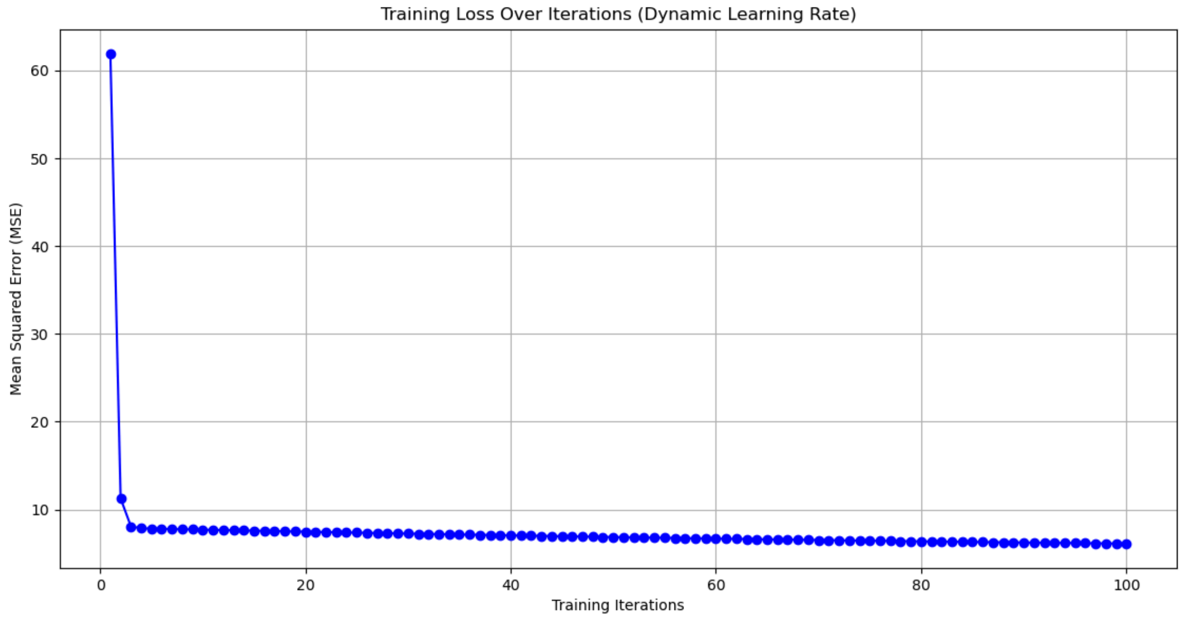
a. α = 0.02, Decay = 0.001, Decay Step = 10

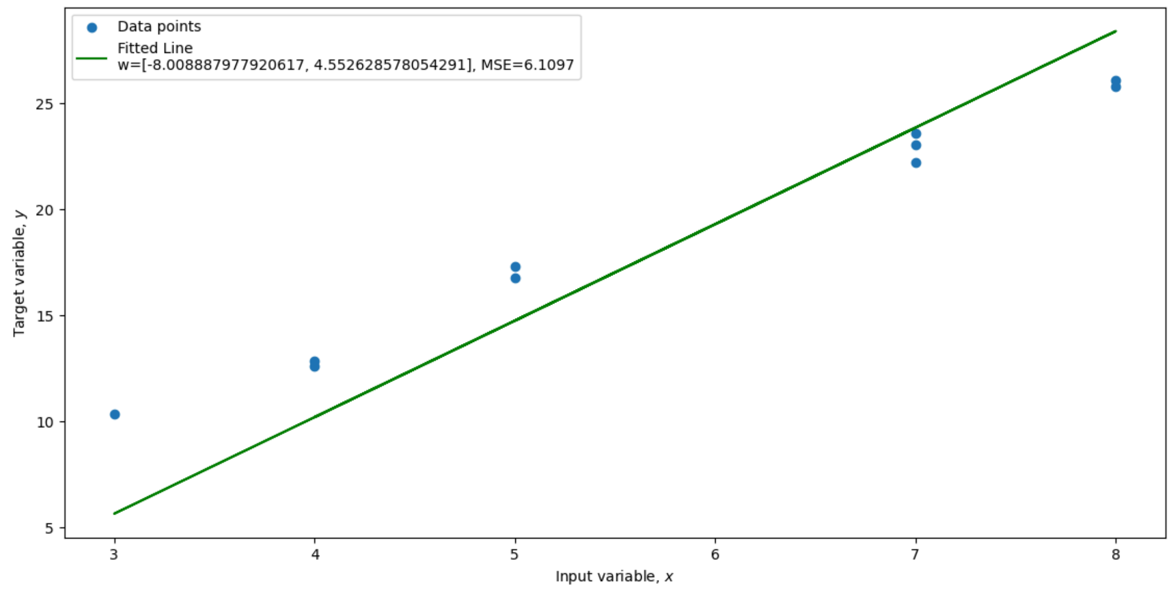
- **Behavior:** Smooth and efficient convergence.
- **Performance:**
  - MSE starts at 225.5152.
  - Rapidly decreases to 7 within 6 iterations.
  - Ends at the lowest MSE of roughly 4.7.
  - The regression line fits the data well.



b. α = 0.01, Decay = 0.001, Decay Step = 10

- **Behavior:** relatively slower convergence than the previous dynamic rate but efficient.
- **Performance:**
  - MSE starts at 61.9519.
  - Decreases to 7.85 in 4 iterations.
  - Ends at 6.1.
  - The regression line fits the data well with a balanced convergence rate.





5. Conclusion

- Lower learning rates (α = 0.001) result in slower but steady convergence.
- Moderate rates (α = 0.01) achieve faster convergence with good stability.
- High learning rates (α = 0.1) cause divergence due to overly large updates.
- **Dynamic learning rates** provide smoother and faster convergence by adjusting the step size over iterations.
- **Observation:** When the optimal fixed learning rate is chosen, the performance is comparable to dynamic learning rate strategies. However, dynamic learning rates offer more stability and adaptability.

Limitations: These conclusions are based on finite testing with a specific dataset and model configuration. Results may vary with different datasets or model architectures.

---

6. Visualizations

- MSE vs. Iterations for each learning rate.
- Fitted regression lines compared with data points.