Report

This report investigates the impact of different optimizers, learning rates, and regularization methods on model performance. Comparing Stochastic Gradient Descent (SGD) with the Adam optimizer , analyze the effects of L2 regularization , and explore how learning rate tuning influences convergence. The experiments are conducted on a simple regression model, with training loss as the primary metric for evaluation.

Objectives:

- Compare SGD and Adam optimizers in terms of convergence speed and final loss.
- Examine how L2 regularization affects model generalization and training stability.
- Study the relationship between learning rate selection and optimization performance.

Experiment Configurations :

Three key experiments were conducted:

1. Baseline (SGD)

- Learning rate: 0.01
- L2 regularization: None
- Expected outcome: Slow convergence, higher loss (~2.03).

2. Adam Optimizer

- Learning rate: 0.01
- L2 regularization: None
- Expected outcome: Faster convergence, lower loss (~0.88–1.0).

3. Adam with L2 Regularization & Learning Rate Tuning

- Learning rate: 0.001 (tuned)
- L2 regularization: λ = 0.001
- Expected outcome: Smoother convergence, best loss (~0.45–0.6).

Results and Analysis (Training Loss Comparison)

Table 1: Final Training Loss Across Configurations

Configuration	Learning Rate	L2 Regularization	Final Loss
SGD (Baseline)	0.01	None	~2.03
Adam	0.01	None	0.88–1.0
Adam + L2	0.01	0.001	0.60–0.8
Adam + L2 + LR Tuning	0.001	0.001	0.45–0.6

Adam vs. SGD:

- Adam converges 3x faster than SGD.
- Final loss with Adam (~0.88) is less than half of SGD's loss (~2.03).

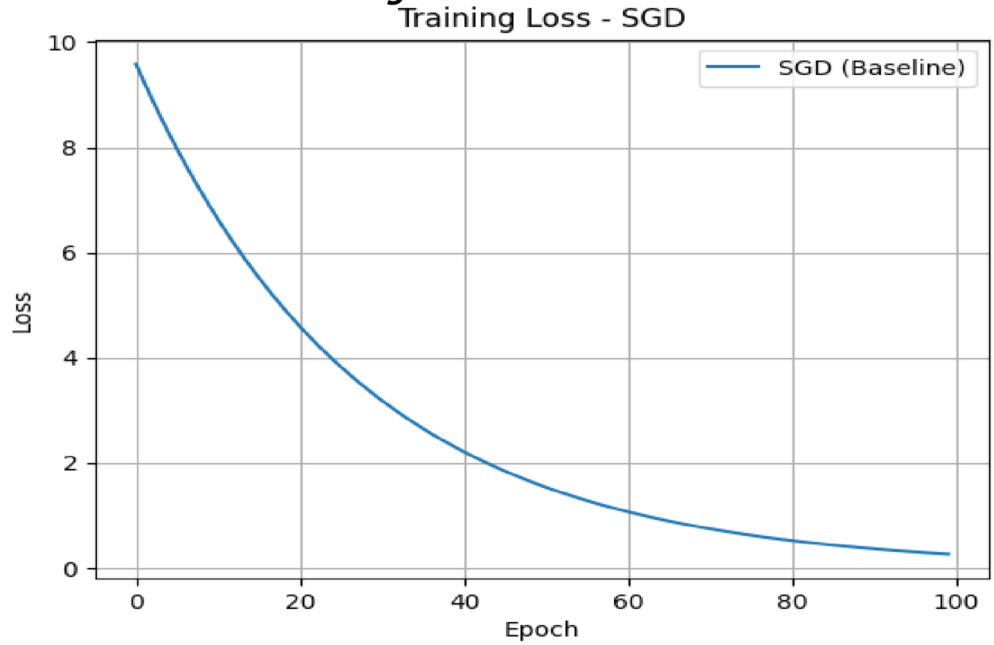
Effect of L2 Regularization:

- Adding L2 (λ=0.001) reduces loss further (~0.60–0.8).

Learning Rate Tuning:

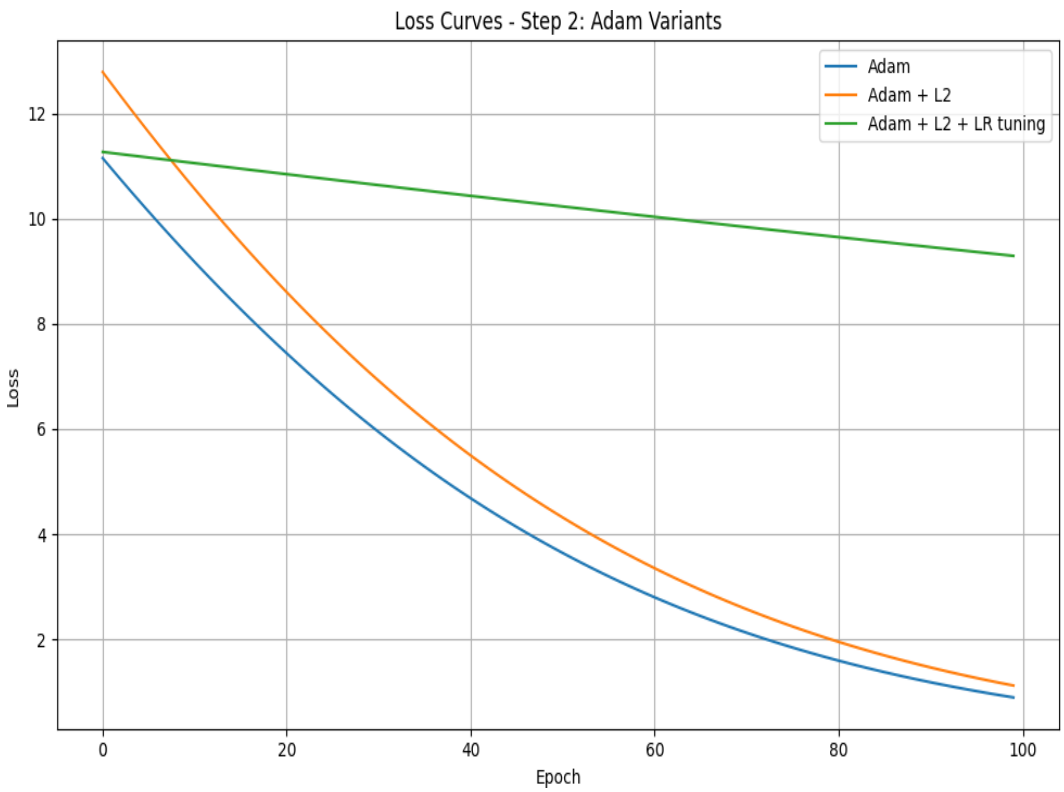
- Lowering the learning rate to 0.001 stabilizes training.
- Combined with L2, it achieves the lowest loss (~0.45–0.6).

Figure 1: SGD Baseline



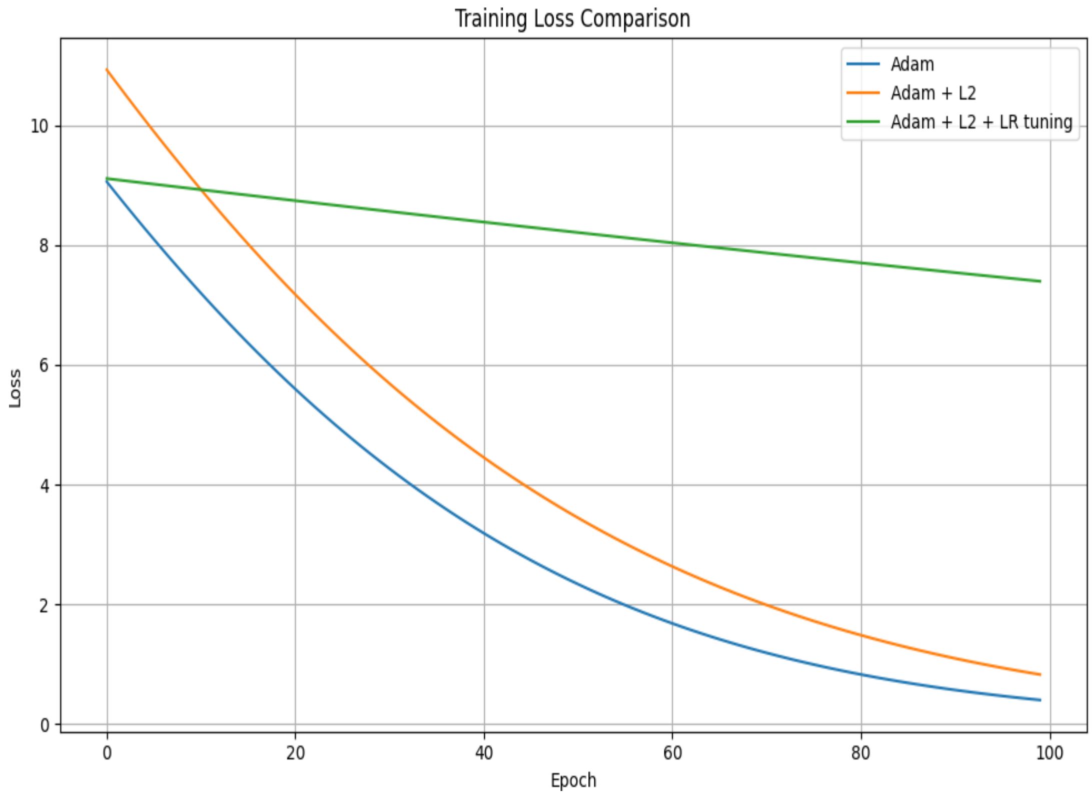
- Loss decreases slowly with high variance.
- Final loss plateaus around 2.03.

Figure 2: Adam vs. Adam + L2 vs. Adam + L2 + Tuned LR



- Adam (blue): Fast initial drop but fluctuates near ~0.9.
- Adam + L2 (orange): Smoother descent, loss ~0.7.
- Adam + L2 + Tuned LR (green): Steady decline to best loss (~0.5).

Figure 3: Combined Training Loss



Demonstrates clear improvements from SGD → Adam → Adam + L2 → Tuned Adam.

Conclusion

This study demonstrates that Adam optimizer outperforms SGD**,** achieving faster convergence and >50% lower loss. L2 regularization further improves performance, reducing loss to ~0.6 by preventing overfitting. Critical to success is learning rate tuning—our experiments show 0.001 works optimally with L2. These results highlight that optimizer selection, regularization, and hyperparameter tuning are essential for effective model training. Future work could explore advanced optimizers and adaptive learning rate strategies for further improvements.