MinLLaMA Assignment 1 Report on Weight Decay

Isaac To

January 29, 2026

1 Experiment Summary

All commands provided in the README and all sanity checks were executed without issue.

2 Weight Decay Optimization

Optimize weight decay parameter to maximize test set accuracy on CFIMDB dataset.

2.1 Methodology

Binary search from 0.0 to 1.0. All experiments used:

- Epochs: 5
- Learning rate: 2e-5
- Batch size: 10
- Dataset: CFIMDB

2.2 Binary Search Results

Weight Decay	Dev Acc	Test Acc
0.01 (baseline)	0.861	0.543
0.25	0.865	0.557
0.5	0.865	0.557
0.75	0.865	0.572
0.875	**0.857**	**0.580**
0.9375	0.849	0.578

Table 1: Weight decay hyperparameter search results

2.3 Key Findings

1. **Optimal Weight Decay:** 0.875 achieved the best test accuracy of **58.0%**
2. **Improvement:** +3.7% absolute improvement over baseline (0.543 → 0.580)
3. **Trend:** Higher weight decay values (0.75-0.875) significantly improved generalization

2.4 Analysis

The improvement likely comes from the fact that we have both a small a model and a small fine-tuning dataset. This means that there is a high chance that we overfit while training. By raising the weight decay this high, we introduce an extremely strong regularization effect forcing the model to learn key connections while dropping superficial connections between items. The main trade-off being slightly lower dev accuracy (-0.4%) for significantly better test accuracy (+3.7%)

3 Full Metrics

Numbers from across all runs

Experiment	**Mode**	**Ep.**	**LR**	**Batch**	**WD**	**Dev**	**Test**
Text Generation	generate	—	—	—	—	—	—
Zero-Shot SST	prompt	—	—	10	—	0.215	0.224
Zero-Shot CFIMDB	prompt	—	—	10	—	0.502	0.213
Fine-Tune SST	finetune	5	2e-5	80	0.01	0.401	0.402
Fine-Tune CFIMDB	finetune	5	2e-5	10	0.01	0.861	0.543
WD Search 0.25	finetune	5	2e-5	10	0.25	0.865	0.557
WD Search 0.50	finetune	5	2e-5	10	0.50	0.865	0.557
WD Search 0.75	finetune	5	2e-5	10	0.75	0.865	0.572
WD Search 0.875	finetune	5	2e-5	10	**0.875**	0.857	**0.580**
WD Search 0.9375	finetune	5	2e-5	10	0.9375	0.849	0.578

Table 2: Command breakdown: Ep. (epochs), WD (weight decay), Dev/Test (accuracy).