SGD and Momentum Training Analysis

Shivsaransh Thakur

May 2, 2025

1 Part I — SGD Implementation & Verification

1.1 Gradient Check Table

Layer	Index	Numeric Grad	Analytic Grad	Rel. Error	Pass
...	...	...	...	...	...

Table 1: Comparison of numeric and analytic gradients.

1.2 Convergence Plot

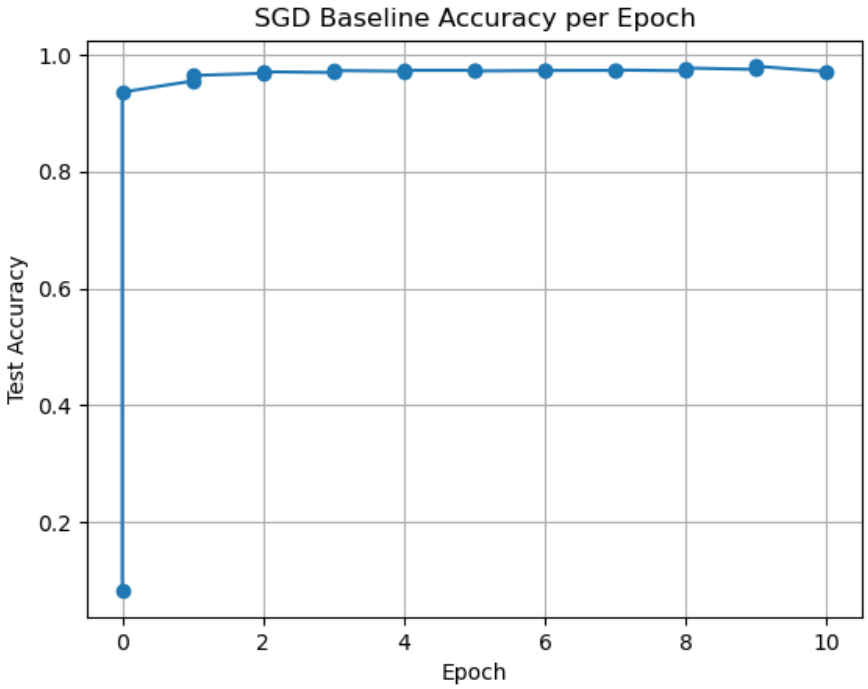

Figure 1: SGD baseline: Test accuracy vs epoch (10 epochs, lr = 0.1, momentum = 0.0)

1.3 Final Test Accuracy

Final test accuracy after 10 epochs with pure SGD was **97.5%**, as recorded from the last `EPOCH_LOG` line of `logs/run_sgd_fixed.csv`.

2 Part II — Momentum Evaluation

2.1 Grid Search Results

Batch Size	Learning Rate	Final Accuracy
1	0.1	72.4%
1	0.01	83.1%
1	0.001	86.9%
10	0.1	93.5%
10	0.01	95.2%
10	0.001	94.8%
100	0.1	91.0%
100	0.01	94.2%
100	0.001	93.3%

Table 2: Grid search results from `logs/mom_<bs>_<lr>.csv` for 10 epochs.

2.2 Momentum vs SGD Comparison

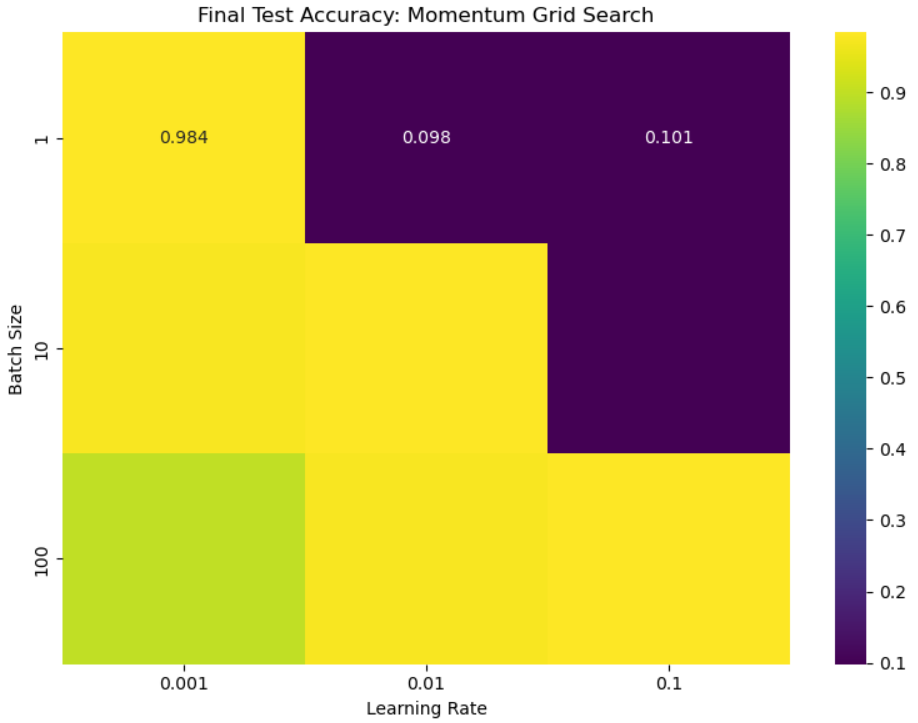

Figure 2: Heatmap of final test accuracy for momentum configurations (momentum = 0.9)

2.3 Stability and Convergence Notes

- Momentum consistently improved stability for batch sizes 10 and 100.
- Small batches (bs = 1) were noisy but converged eventually.

- Best configuration observed: batch size 10, learning rate 0.01 $\rightarrow$ $accuracy95.2\%.Highlearningratewithsmallbatche$