

WILLIAM BLACKWOOD, Department of Computer Science and Mathematics, Fairmont State University, 1201 Locust Ave, Fairmont, WV 26554. An empirical study of nonconvex optimization in deep learning.

This study examines how common training choices affect convergence and generalization in nonconvex deep learning. Models were trained on CIFAR-10 using ResNet-18, VGG-11, and a three-layer MLP while varying optimizer, learning-rate schedule, batch size, and weight decay. Each configuration was repeated across three random seeds and evaluated using training loss, test accuracy, generalization gap, and an estimate of Hessian trace as a measure of sharpness. The results show that adaptive optimizers reduced loss faster early in training, but SGD with momentum remained competitive in final test accuracy when properly tuned. Cosine annealing and warmup plus cosine schedules generally produced smaller generalization gaps than constant and step schedules. Increasing batch size was associated with sharper minima and worse generalization, while AdamW produced more reliable regularization than standard Adam under matched weight-decay settings. Overall, the study shows that optimizer and training choices strongly influence both convergence behavior and final model quality, and that careful tuning can matter as much as the optimizer itself.