Date

14/18 June 2023

Topic: Sharpness - Aware minimalization for efficiency improving generalization

Abstract:

In today's heavily overparameterized models, the value of training loss provides few guarantees on model generalization ability. Indeed, optimizing only the training loss value, as is commonly done, can easily lead to suboptimal model quality. Motivated by prior work connecting the geometry of the loss landscape and generalization, we introduce a novel, effective procedure for instead simultaneously minimizing loss value and loss sharpness. In particular our procedure, Sharpness-Aware Minimization (SAM) seeks parameters that lie in neighborhoods having uniformly low loss, this formulation results in a min max optimization problem on which gradient descent can be performed efficiently. We present empirical results showing that SAM improves model generalization across a ~~variety~~ variety of benchmark datasets and models, yielding novel state of the art performance for several. Additionally, we find that SAM natively provides robustness to label noise with that provided by state of the art procedures

that specifically target learning with noisy labels
Conclusion
In this work we have introduced SAM a novel algorithm that improves generalization by simultaneously by minimizing loss value and loss sharpness, we have demonstrated SAM's efficacy through a rigorous large scale empirical evaluation We have surfaced a number of interesting avenues for future work On the theoretical ~~work~~ side, the notion of per data point sharpness yielded by m-sharpness suggests an interesting new lens through which to study generalization. Methodologically, our methods that currently rely on mixup. We leave to future work for more in depth investigation of these possibilities.