Computer Science > Machine Learning

arXiv:1803.09820 (cs)

[Submitted on 26 Mar 2018 (v1), last revised 24 Apr 2018 (this version, v2)]

A disciplined approach to neural network hyper-parameters: Part 1 -- learning rate, batch size, momentum, and weight decay

Leslie N. Smith

Although deep learning has produced dazzling successes for applications of image, speech, and video processing in the past few years, most trainings are with suboptimal hyper-parameters, requiring unnecessarily long training times. Setting the hyper-parameters remains a black art that requires years of experience to acquire. This report proposes several efficient ways to set the hyper-parameters that significantly reduce training time and improves performance. Specifically, this report shows how to examine the training validation/test loss function for subtle clues of underfitting and overfitting and suggests guidelines for moving toward the optimal balance point. Then it discusses how to increase/decrease the learning rate/momentum to speed up training. Our experiments show that it is crucial to balance every manner of regularization for each dataset and architecture. Weight decay is used as a sample regularizer to show how its optimal value is tightly coupled with the learning rates and momentums. Files to help replicate the results reported here are available.

Comments: Files to help replicate the results reported here are available on Github
Subjects: **Machine Learning (cs.LG)**; Computer Vision and Pattern Recognition (cs.CV); Neural and Evolutionary Computing (cs.NE); Machine Learning (stat.ML)
Report number: US Naval Research Laboratory Technical Report 5510-026
Cite as: arXiv:1803.09820 [cs.LG]
(or arXiv:1803.09820v2 [cs.LG] for this version)
https://doi.org/10.48550/arXiv.1803.09820
Focus to learn more
arXiv-issued DOI via DataCite

Submission history

From: Leslie Smith [view email]
[v1] Mon, 26 Mar 2018 20:05:59 UTC (3,871 KB)
[v2] Tue, 24 Apr 2018 17:43:51 UTC (3,871 KB)

Core recommender toggle
CORE Recommender *(What is CORE?)*
IArxiv recommender toggle
IArxiv Recommender *(What is IArxiv?)*

About arXivLabs

arXivLabs: experimental projects with community collaborators

arXivLabs is a framework that allows collaborators to develop and share new arXiv features directly on our website.

Both individuals and organizations that work with arXivLabs have embraced and accepted our values of openness, community, excellence, and user data privacy. arXiv is committed to these values and only works with partners that adhere to them.

Have an idea for a project that will add value for arXiv's community? **Learn more about arXivLabs** and **how to get involved**.