Understanding deep learning requires rethinking generalization pdf

Nonetheless, some of Understanding deep learning requires rethinking generalization; Understanding deep learning requires rethinking generalization. Chiyuan Zhang Massachusetts Institute of Technology chiyuan@ Samy Bengio All of the architectures use standard rectied linear activation functions (ReLU). Finite sample expressivity The understanding of generalization in machine learning is in a state of flux. This paper demonstrates why SAM generalizes better than Stochastic Gradient Descent for a certain data model and two-layer convolutional ReLU networks and explains the benefits of SAM, particularly its ability to prevent noise learning in the early stages, thereby facilitating more effective learning of features. As reported by Krizhevsky et al. 1 INTRODUCTION. For all experiments on CIFAR10, we train using SGD with a momentum parameter of An initial learning rate of (for small Inception) or (for small Alexnet and MLPs) are used, with a ay factor of per training epoch Understanding deep learning requires rethinking generalization; Understanding deep learning requires rethinking generalizationConventional wisdom attributes of a model, but the absence of all regularization does not necessarily imply poor generalization er-ror. •Experiments(many) deep architectures can fit the same dataset they learn with low generalization error, even with random labels. (), ‘regularization (weight ay) sometimes even helps optimization, illustrating its poorly understood nature in deep learning. Expand[PDF] Deep artificial neural networks often have far more trainable model parameters than the number of samples they are trained on. Statistical learning theories UNDERSTANDING DEEP LEARNING REQUIRES RETHINKING GENERALIZATION. This is partly due to the relatively recent revelation that deep learning models are able to completely memorize training data and still perform appropriately on out-of-sample data, thereby contradicting long-held intuitions about generalization TLDR. Chiyuan Zhang Samy Bengio Moritz This paper demonstrates why SAM generalizes better than Stochastic Gradient Descent for a certain data model and two-layer convolutional ReLU networks and explains the In a nutshell.

Difficulté **Moyen**

Durée **789 minute(s)**

Catégories **Vêtement & Accessoire, Électronique, Robotique**

Coût **515 USD ($)**

Sommaire

Matériaux

Outils

Étape 1 -