Overview of Robust Fine-Tuning

Week 37

202111309 손본

Abstract

해당 분야에서 주로 등장하는 용어에 대한 정리, 앞으로 관련 논문을 읽기 위하여 공부해야할 내용 정리

Robust Fine-Tuning of Deep Neural Networks with Hessian-based Generalization Guarantees

Abstract

We study the generalization properties of fine-tuning to understand the problem of overfitting, which has often been observe.

- target dataset is small
- training labels are noisy

Existing generalization measures for deep networks depend on notions such as distance from the initialization (i.e., the pretrained network) of the fine-tuned model and noise stability properties of deep networks.

- distance : 모델의 초기(=pretrained network)로부터 fine-tuning을 끝낸 모델의 차이
- noise stability : Adversarial attack 등에 얼마나 강건성을 보이는가

-> 위의 2가지를 일반적으로 현재 신경망이 얼마나 잘 일반화되어있는가를 판단하는 지표로서 많이 사용한다.

Abstract

This paper identifies a Hessian-based distance measure through PAC-Bayesian analysis, which is shown to correlate well with observed generalization gaps of fine-tuned models

prove **Hessian distance-based generalization bounds for fine-tuned model**

- We present an algorithm and a generalization error guarantee for this algorithm under a class conditional independent noise mode
- Hessian-based distance measure can match the scale of the observed generalization gap of fine-tuned models in practice.

Introduction

fine-tuning

fine-tuning is common method to using large pre-trained model into real probblem.

Understanding the cause of overfitting is challenging since dissecting the issue in practice requires a precise measurement of generalization errors for deep neural networks

Introduction

In this work,

In this work, we analyze the generalization error of fine-tuned deep models using PAC-Bayes analysis and data-dependent measurements based on deep net Hessian. With this analysis, we also study the robustness of fine-tuning against label noise

> **Note**
>
> There is a large body of work concerning generalization in deep neural networks, **whereas less is understood for fine-tuning**

Introduction

선행 연구

These results highlight that **distance from initialization** crucially affects generalization for fine-tuning and informs the design of distance-based regularization to mitigate overfitting due to fine-tuning a large model on a small training set

1. 초기 가중치로부터의 거리가 일반화에 결정적으로 영향을 미친다.
2. 거리 기반의 정규화를 사용하여 과적합을 완화할 수 있다.

> **Info**
>
> distance-based regularization을 이용하여 small train set에 대해 큰 모델을 fine-tunning할 때 발생하는 overfitting 문제를 완화할 수 있다.

Introduction

아이디어 및 접근법들

1. PAC-Bayesian analysis approach
2. deep net Hessian

> **Idea**
>
> We propose to quantify the stability of a deep model against noise perturbations using Hessian

헤시안을 사용하여 Deep net의 perturbation에 대한 안정성을 정량화하려고 한다.

Introduction

Hessian distance measure that better captures emirical generalization erros of fine-tuned models.

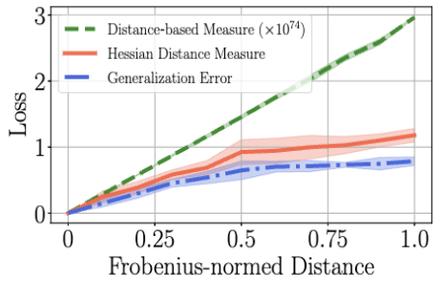


Figure 1: We identify a Hessian distance measure (cf. equation (1)) that better captures empirical generalization errors of fine-tuned models.

Method	Generalization error bound
Fine-tuning	$\frac{\sum_{i=1}^{L}\sqrt{v_i^\top \mathbf{H}_i^+ v_i}}{\sqrt{n}}$
Distance-based Reg.	$\frac{\sum_{i=1}^{L}\sqrt{\mathrm{Tr}\left[\mathbf{H}_i^+\right]\cdot\|v_i\|^2}}{\sqrt{n}}$
Consistent loss w/ Reg.	$\frac{\sum_{i=1}^{L}\sqrt{\left\|(F^{-1})^\top\right\|_{1,\infty}\cdot\|\mathrm{Tr}[\mathbf{H}_i]\|\cdot\|v_i\|^2}}{\sqrt{n}}$

Table 1: A summary of the theoretical bounds: To use these results in practice, we can compute Hessian-vector product libraries. See Section 2.1 for the definition of these notations.

=> 즉, 논문에서는 모델이 얼마나 잘 일반화되는지를 이해하기 위해 initial weight에서 얼마나 멀리 떨어져 있는지(=distance)와 Hessian을 모두 고려하여 평가하는 방법을 제안한다. 특히 이는 fine-tuning 과정에서 모델의 일반화 능력을 정확하게 예측하는 데 도움이 될 수 있다.

Terminology

Terminology

generalization gap

보통 모델의 training accuracy와 test accuray사이의 차이로, 모델의 일반화 성능을 나타내는 지표로서 사용된다.

noisy labels

Learning with noisy labels means When we say **"noisy labels,"** we mean that an adversary has intentionally messed up the labels, which would have come from a "clean" distribution otherwise. This setting can also be used to cast learning from only positive and unlabeled data.

PAC-Bayes analysis

Bayesian learning에서 사용되었으며 현재는 일반적인 상황에도 적용가능하다.
PAC-Bayes theory gives the tightest known generalization bounds for SVMs !

Study : leacture_PAC-Bayes analysis, An Introduction to PAC-Bayesian Analysis, ICML 2021 튜토리얼

Terminology

Hessian

어떤 함수의 이계도함수 를 행렬로 표현한 것
실함수 $f(x_1, x_2, x_3, \cdots, x_n)$이 주어졌을 때, **Hessian Matrix**은 다음과 같이 주어진다.

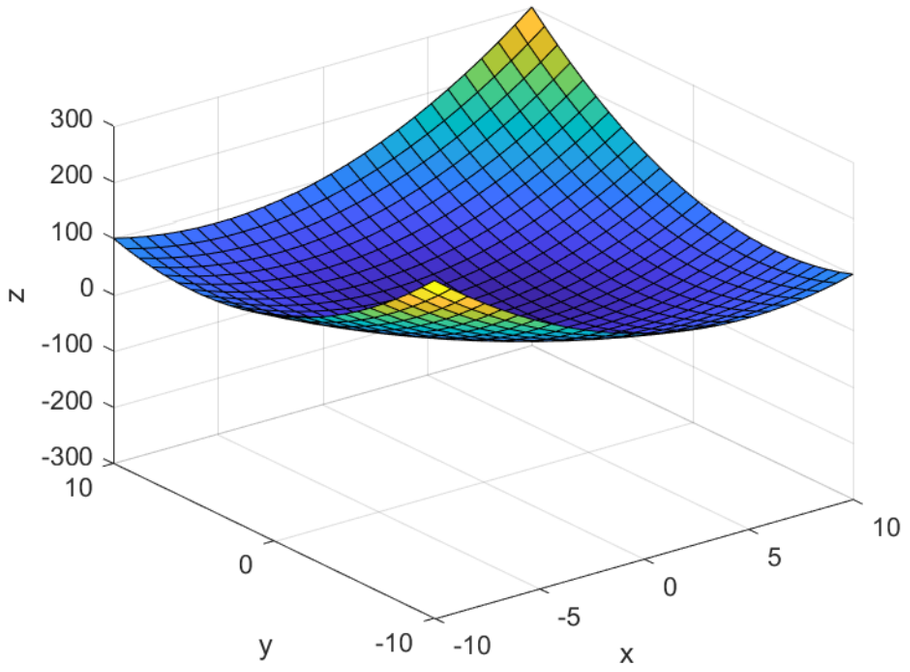
즉, Hessian Matrix의 H_{ij} 요소는 x_i와 x_j 변수에 대해 함수를 이차미분한 값이다.

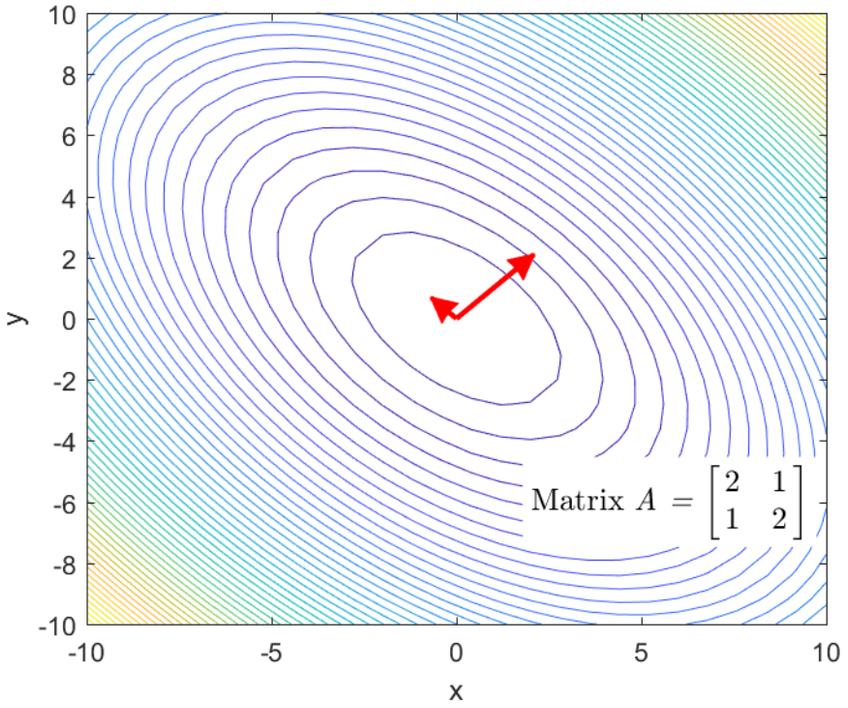
$$H(f) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

이때, 함수 f의 이계도함수가 연속이라면 혼합 편미분은 같기 때문에 f의 이계도함수가 연속이라면 헤세 행렬은 대칭행렬(symmetric matrix)이다.

Terminology

헤시안의 의미

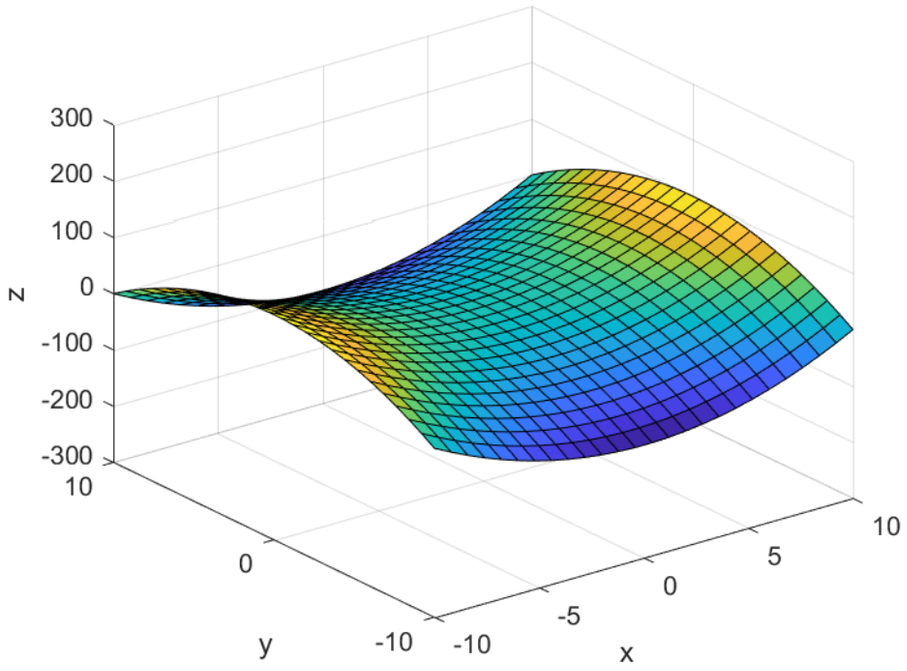


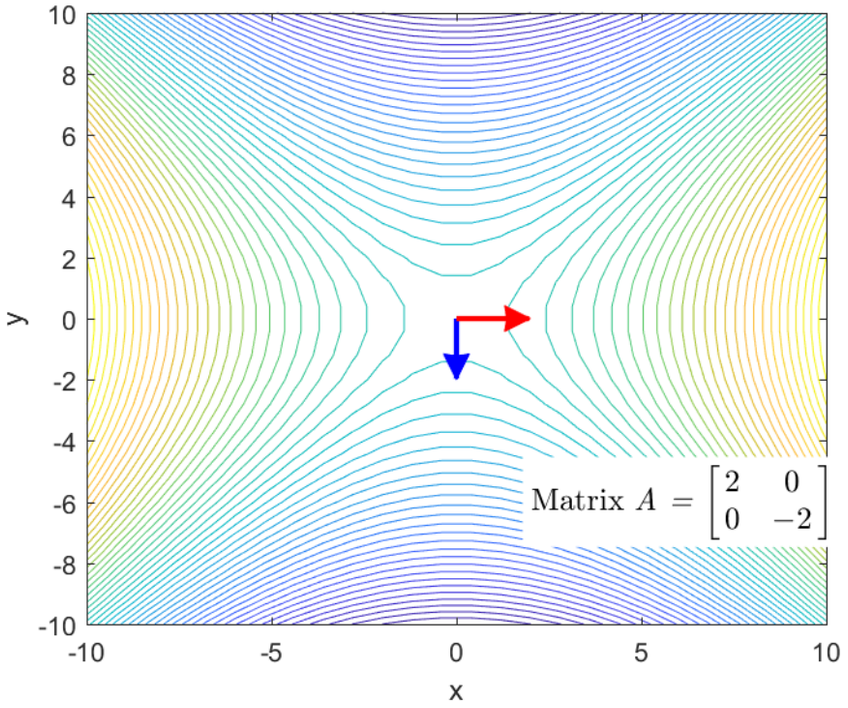


Reference : 헤세 행렬의 기하학적 의미

Terminology

헤시안의 의미





Terminology

헤시안의 의미

- 특정 고유벡터에 대해 고유값의 절댓값의 크기가 클수록 해당 방향으로 더 가파르게 변화한다.
- **헤시안 행렬의 고윳값이 모두 양수라면 함수는 아래로 볼록하다. 일차미분이 0이라면 극솟값** ; convex
- 헤시안 행렬의 고윳값이 모두 음수라면 함수는 위로 볼록하다. 일차미분이 0이라면 극댓값
- 헤시안 행렬의 교윳값에 양수와 음수가 섞여있는 경우라면 함수는 안장의 형태. 일차미분이 0이라면 안장점

해시안의 의미 (한줄 정리)

"헤시안(Hessian)"은 머신러닝과 최적화에서 중요한 역할을 하는 수학적 개념으로, **주어진 함수의 곡률**을 설명하기 위하여 사용된다. 헤시안은 모델(함수)이 가지는 매개변수 공간에서 함숫값이 얼마나 빠르게 변화하는지(즉, 얼마나 '곡률이 높은지' 또는 '곡률이 낮은지')를 알려주는 역할을 수행한다.

Hessian 행렬은 Convex Optimization, 이계도함수 판정, Newton's Method 등의 여러가지 방법에 활용된다.

Paper

읽어볼 논문 들 ...

Paper

B. Neyshabur, H. Sedghi, and C. Zhang. “What is being transferred in transfer learning?” In: NeurIPS (2020) : 전이학습이 왜 가능한지, 어떤 요소가 크게 영향을 미치는지, 전이 학습에서 찾을 수 있는 Optimal point는 어디인지에 대한 연구 <291>

P. Bartlett, D. J. Foster, and M. Telgarsky. “Spectrally-normalized margin bounds for neural networks”. In: Neural Information Processing Systems (2017) <1085> : margin bound에 대한 연구 (1)

B. Neyshabur, S. Bhojanapalli, and N. Srebro. “A pac-bayesian approach to spectrally-normalized margin bounds for neural networks”. In: ICLR (2018) <590> : margin bound에 대한 연구 (2)

V. Nagarajan and J. Z. Kolter. “Generalization in deep networks: The role of distance from initialization”. In: arXiv preprint arXiv:1901.01672 (2019) <76> : generalization 연구

P. M. Long and H. Sedghi. “Generalization bounds for deep convolutional neural networks”. In: ICLR (2020) <62> : generalization 연구

D. Li and H. Zhang. “Improved Regularization and Robustness for Fine-tuning in Neural Networks”. In: NeurIPS (2021) : fine-tuning을 더욱 효과적으로 하고, 그 과정에서 발생할 수 있는 과적합과 노이즈 문제를 해결하기 위한 새로운 방법론을 제안 (with few-shot learning) <26>

Paper

논문에서 핵심적으로 언급한 기술에 대한 논문

PAC-Bayesian analysis approach

- B. Neyshabur, S. Bhojanapalli, and N. Srebro. “A pac-bayesian approach to spectrally- normalized margin bounds for neural networks”. In: ICLR (2018)
- S. Arora, R. Ge, B. Neyshabur, and Y. Zhang. “Stronger generalization bounds for deep nets via a compression approach”. In: International Conference on Machine Learning. 2018

deep net Hessian

- Z. Yao, A. Gholami, K. Keutzer, and M. W. Mahoney. “Pyhessian: Neural networks through the lens of the hessian”. In: IEEE International Conference on Big Data. 2020

PYHESSIAN은 딥 러닝 신경망에 대한 헤시안(즉, 2차 미분) 정보의 빠른 계산을 가능하게 하는 프레임 워크이다. 이를 이용하면 신경망의 loss landscape의 토폴로지를 이해하기 위한 정보를 얻을 수 있다. (곡률)