

深層ニューラルネットワークにおける二重降下現象

Double Descent in Deep Neural Networks

高橋秀弥¹
Shuya Takahashi

鏡川悠介¹
Yusuke Kyokawa

前田英作¹
Eisaku Maeda

¹ 東京電機大学

1 はじめに

近年, Belkin らは, モデルサイズを増加させていくにつれて検証用データに対する不正解率が減少し, 増大, また減少に転じる Double Decent (DD) という現象が起きることを示した [1]. Nakkiran らは複数の深層ニューラルネットワークで DD が起こることを示した. また, 従来の DD を Model-wise DD (MDD) とし, 学習の進みに対して起こる Epoch-wise DD (EDD) を観測した [2]. 我々は, 二重降下現象における仮説検証を行うために Nakkira らの手法に関する再現実験, 条件を変更した場合の EDD 観測実験を実施した.

2 実験

Nakkiran らの実験方法を参考に MDD, EDD の再現実験を行うとともに, 事前学習 (PT) の有無で EDD にどのような影響があるかを比較するための追加実験を行った. 学習データセットには CIFAR-10 (C10) を用意した. C10 は訓練用画像を 5 万枚, 検証用画像を 1 万枚もつ自然画像データセットである. 二重降下の再現にあたり, 訓練データの正解ラベルを $p\%$ の割合で異なったラベルに張り替える操作を各実験で行った. データ拡張にはランダムに画像を切り出す RandomCrop, 画像をランダムに左右反転する RandomHorizontalFlip を実施した. 再現実験では学習モデルに ResNet18 を横幅可変にした ResNet18k¹ を使用した. ResNet18k の横幅に 64 を指定すると各層の出力チャンネルは ResNet18 と同一になる. 損失関数は交差エントロピー誤差を, 最適化関数には Adam を使用し, 学習率を 0.0001 に設定した.

MDD の再現実験では C10 データセットに 15% のラベルノイズを付与し, ResNet18k の横幅を 1 から 64 まで変化させ, 各モデルを 4000epoch まで学習を行った. モデルの複雑さに応じて不正解率が二重降下するか観測した.

EDD の再現実験では C10 データセットに 20% のラベルノイズの付与し, 横幅 128 の ResNet18k を 1500 epoch 学習した. モデルの学習回数に応じて不正解率が二重降下するかを観測した.

追加の実験では ImageNet での PT を行った ResNet18 と行っていない ResNet18 を用いてそれぞれ 4000 epoch 学習し, 二重降下の有無とその影響を観測した.

3 結果

モデルの複雑性に応じて不正解率が二重降下する MDD の結果を図 1, 学習回数に応じて不正解率が二重降下す

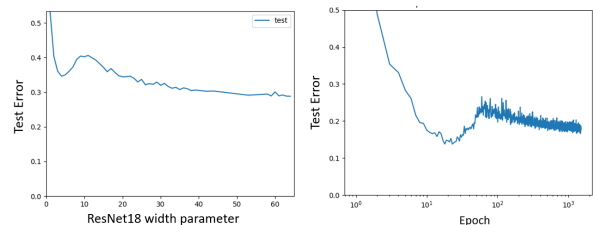


図 1: Model-wise DD

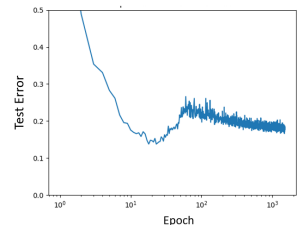


図 2: Epoch-wise DD

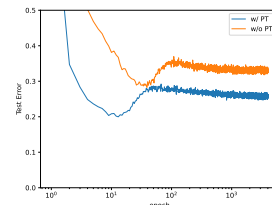


図 3: EDD w/ w/o pre-train

る EDD の結果を図 2 に, 追加の実験結果を図 3 に示す. 実験結果から二重降下現象を再現することができた. また, ResNet18 及び PT ありという条件でも同様に二重降下を観測できた.

4 考察

PT の有無で EDD を比較すると PT ありのものはより早く Test Error (TE) の下降が終わり上昇に転じている. これは PT なしのものが何も学習していない状態から学習しているのに対して事前学習ありのものはある程度自然画像というものを学習した状態から C10 を学習するため, より早く最適解にたどり着くからだと考える. また, PT なしのものの TE はこのまま学習を続けても PT ありの TE に到達することはないように見受けられる. この理由として PT することで IN からより豊かな特徴を学習し, より高い性能を得られる最適解の付近から学習を始められるが, PT なしの場合まったくのランダムなパラメータから学習を始めることで局所解にはまってしまうからだと考える.

5 おわりに

本稿では MDD, EDD ともに二重降下現象を再現した. 今後は C10 に限らないデータセットで二重降下現象を確認することや特徴の複雑性に関する検証を行う.

参考文献

- [1] M. Belkin *et al.*, “Reconciling modern machine-learning practice and the classical bias–variance trade-off,” *PNAS*, vol. 116, no. 32, pp. 15 849–15 854, 2019.
- [2] P. Nakkiran *et al.*, “Deep double descent: where bigger models and more data hurt,” in *ICLR*, 2020.

¹<https://gitlab.com/harvard-machine-learning/double-descent/-/blob/master/models/resnet18k.py>