Commun. Comput. Phys.
doi: 10.4208/cicp.OA-2021-0211

Vol. **32**, No. 4, pp. 1007-1038
October 2022

Generalization Error Analysis of Neural Networks with Gradient Based Regularization

Lingfeng Li[1], Xue-Cheng Tai[1,*] and Jiang Yang[2,3]

[1] *Hong Kong Centre for Cerebro-Cardiovascular Health Engineering, 19W, Hong Kong Science Park, Shatin, Hong Kong, China.*
[2] *Department of Mathematics, Southern University of Science and Technology, Shenzhen, China.*
[3] *SUSTech International Center for Mathematics, Southern University of Science and Technology, Shenzhen, China.*

Received 27 October 2021; Accepted (in revised version) 22 July 2022

Abstract. In this work, we study gradient-based regularization methods for neural networks. We mainly focus on two regularization methods: the total variation and the Tikhonov regularization. Adding the regularization term to the training loss is equivalent to using neural networks to solve some variational problems, mostly in high dimensions in practical applications. We introduce a general framework to analyze the error between neural network solutions and true solutions to variational problems. The error consists of three parts: the approximation errors of neural networks, the quadrature errors of numerical integration, and the optimization error. We also apply the proposed framework to two-layer networks to derive a priori error estimate when the true solution belongs to the so-called Barron space. Moreover, we conduct some numerical experiments to show that neural networks can solve corresponding variational problems sufficiently well. The networks with gradient-based regularization are much more robust in image applications.

AMS subject classifications: 68T07

Key words: Machine learning, regularization, generalization error, image classification.

1 Introduction

Deep neural networks (DNNs), which are compositions of some linear and non-linear mappings, have become the most popular tool in artificial intelligence. Its ability to fit complex functions helps it achieve state-of-the-art performances and beat other methods

*Corresponding author. *Email addresses:* lfli@hkcoche.org (L. Li), xtai@hkcoche.org (X.-C. Tai), yangj7@sustech.edu.cn (J. Yang)

by a huge margin in many areas, such as image processing, video processing, and natural language processing. For a complete introduction to deep learning, one can refer to [1–3]. Recently, some applied mathematicians have also successfully applied DNNs to solve some partial differential equations (PDEs) [4–10]. The main advantage of DNNs is that they can solve very high-dimensional problems that are intractable for traditional numerical methods. Some literature about the expressive power of DNNs are [11–14].

Though DNNs have such great expressive power, the training of DNNs is not easy, especially for those having enormous amounts of parameters. One of the most common issues in training DNNs is overfitting, i.e., the model fits the training data well but performs poorly on the testing data. Overfitting is very likely to happen when the number of parameters is significantly larger than the number of training samples. To prevent models from overfitting, some regularization techniques are usually applied.

Two types of regularization methods are commonly used in practice: implicit regularization and explicit regularization. The implicit method is often induced by the optimization scheme and network architecture design. Some popular implicit regularization methods are early stopping, data augmentation, and dropout. Recently, [15,16] proposed the adaptive activation functions which multiplies a learnable parameter with the input of regular activation functions. This new design can dynamically adjust the loss function's landscape and accelerate the convergence by avoiding local minimas. The explicit method is introducing a regularization term directly into the loss function.

Another issue we concern about the DNNs is stability, which is also referred to as adversarial robustness. It has been shown that DNNs can be fooled by adding very small perturbation to the inputs [17, 18]. The algorithms to find such a perturbation is called adversarial attacks, such as FGSM [17], PGD [19], and one-pixel attack [20]. To improve the robustness of our networks, various adversarial defensive strategies have been developed [17, 21], and most of them are in the form of explicit regularization.

Regularization plays important roles not only in deep learning but also in variational models where the regularization terms are often designed based on some prior knowledge [22–24]. One of the most popular regularization methods is the gradient-based regularization like the total variation (TV) [22] and the Tikhonov regularization [25]. The Euler-Lagrangian equations of the regularized problems are some PDEs. Numerical methods like finite difference are commonly used to solve them. However, due to the curse of dimensionality, most of the traditional numerical methods are only able to handle low-dimensional problems. Therefore, using DNNs to approximate solutions to variational models and PDEs have attracted extensive attentions in recent years [4–6, 26–28].

For classical numerical methods, error analysis is one of the most critical parts of the research. Usually, we would expect an explicit bound on the difference between the numerical solution and the true solution. If the error bound depends on the computed numerical solution, we call it a posterior error. Otherwise, we call it a priori error. However, because of the non-convexity of the problems and the randomness induced by the optimization scheme, conducting error analysis for deep learning algorithms is difficult in general. Recently, some generalization error analysis have been developed for PDE-