QIANG YE

Department of Mathematics
University of Kentucky

Batch Normalization and Preconditioning for Neural Network Training

THURSDAY, April 14, 2022 at 4PM
Jones 303, 5747 S. Ellis Ave. Chicago, IL 60637

Batch normalization (BN) is a popular method in deep neural network training that has been shown to decrease training time and improve generalization performance. Despite its success, BN is not theoretically well understood. It is not suitable for use with very small mini-batch sizes or online learning. In this talk, we will review BN and present a preconditioning method called Batch Normalization Preconditioning (BNP) to accelerate neural network training. We will analyze the effects of mini-batch statistics of a hidden variable on the Hessian matrix of a loss function and propose a parameter transformation that is equivalent to normalizing the hidden variables to improve the conditioning of the Hessian. Compared with BN, one benefit of BNP is that it is not constrained on the mini-batch size and works in the online learning setting. We will present several experiments demonstrating competitiveness of BNP. Furthermore, we will discuss a connection to BN which provides theoretical insights on how BN improves training and how BN is applied to special architectures such as convolutional neural networks. The talk is based on a joint work with Susanna Lange and Kyle Helfrich.