

Improvements and extensions to the K-FAC method for training neural nets

James Martens, DeepMind

Second order optimization methods have the potential to be much faster than first order methods in the deterministic case, or pre-asymptotically in the stochastic case. However traditional second order methods have proven ineffective or impractical for neural network training, due in part to the extremely high dimension of the parameter space.

Kronecker-factored Approximate Curvature (K-FAC) is second-order optimization method based on a tractable approximation to the Gauss-Newton/Fisher matrix that exploits the special structure of neural network training objectives. This approximation is neither low-rank nor diagonal, but instead involves Kronecker-products, which allows for efficient estimation, storage and inversion of the curvature matrix.

In this talk I will introduce the basic K-FAC method for standard MLPs and then present some more recent work in this direction, including extensions to different models and/or loss functions such as convnets, RNNs, VAEs, and reinforcement learning agents (each of which requires new approximations and/or formulations of the curvature matrix). I will provide both theoretically motivated arguments for these approximations, as well as empirical results which speak to their efficacy.