Stochastic continuous-depth neural networks

The candidate advances the understanding of deep neural networks through the investigation of stochastic continuous-depth neural networks and introduces novel training algorithms based on recent developments in back-propagation algorithms for stochastic differential equations.

Deep neural networks can represent very complex, nonlinear relationships, while they can still be trained efficiently. But deep neural networks are difficult to reason about mathematically. And while a trained neural network provides a strong predictor, assessing the uncertainty of its predictions is difficult.

Recent papers have drawn a connection between ordinary differential equations and residual neural networks, a type of deep neural networks. See for example Haber and Ruthotto (2018), Chen, Rubanova, Bettencourt, and Duvenaud (2018) or Ruthotto and Haber (2018). In these networks each single network layer only brings about a gradual change to the output of the preceding layers. In the limit of many layers of infinitesimal changes they can be described by a differential equation. Such continuous-depth networks share the expressiveness of DNNs and can be trained efficiently through backpropagation. Their analytical structure also brings new ways to mathematically reason about them.

The efficient training of continuous-depth neural networks from Chen, Rubanova, Bettencourt, and Duvenaud (2018) requires solving an augmented system backwards in time. van der Meulen and Schauer (2017) introduces a corresponding adjoint method for nonlinear stochastic differential equations allowing the candidate to extend the work of Chen, Rubanova, Bettencourt, and Duvenaud (2018) to the continuous-depth neural networks with stochastic regularisation considered in Wang, Yuan, Shi, and Osher (2018). Through the intrinsic noise, such a model is naturally regularised and gives principled uncertainty quantification. The candidate investigates the theoretical and computational merits of this approach.

References

Abarbanel, Henry, Rozdeba, Paul, and Shirman, Sasha (2017). *Machine Learning, Deepest Learning: Statistical Data Assimilation Problems*. eprint: `arXiv:1707.01415`.

Chen, Ricky T. Q., Rubanova, Yulia, Bettencourt, Jesse, and Duvenaud, David (2018). “Neural Ordinary Differential Equations”. In: *arXiv e-prints*. arXiv: `1806.07366`.

Haber, Eldad and Ruthotto, Lars (2018). “Stable architectures for deep neural networks”. In: *Inverse Problems* 34.1, p. 014004.

Ruthotto, Lars and Haber, Eldad (2018). “Deep Neural Networks motivated by Partial Differential Equations”. In: *CoRR* abs/1804.04272. arXiv: `1804.04272`.

van der Meulen, Frank and Schauer, Moritz (2017). *Continuous-discrete smoothing of diffusions*. eprint: `arXiv:1712.03807`.

Wang, Bao, Yuan, Binjie, Shi, Zuoqiang, and Osher, Stanley J. (2018). “EnResNet: ResNet Ensemble via the Feynman-Kac Formalism”. In: *CoRR* abs/1811.10745. arXiv: `1811.10745`.