

CRUNCH Seminars at Brown, Division of Applied Mathematics

Friday – March 20, 2020

The Case for Bayesian Deep Learning

Xuhui Meng

The key distinguishing property of a Bayesian approach is marginalization instead of optimization, not the prior, or Bayes rule. Bayesian inference is especially compelling for deep neural networks. (1) Neural networks are typically underspecified by the data, and can represent many different but high performing models corresponding to different settings of parameters, which is exactly when marginalization will make the biggest difference for both calibration and accuracy. (2) Deep ensembles have been mistaken as competing approaches to Bayesian methods, but can be seen as approximate Bayesian marginalization. (3) The structure of neural networks gives rise to a structured prior in function space, which reflects the inductive biases of neural networks that help them generalize. (4) The observed correlation between parameters in flat regions of the loss and a diversity of solutions that provide good generalization is further conducive to Bayesian marginalization, as flat regions occupy a large volume in a high dimensional space, and each different solution will make a good contribution to a Bayesian model average. (5) Recent practical advances for Bayesian deep learning provide improvements in accuracy and calibration compared to standard training, while retaining scalability.