Baidu: Research Topics

Belhal Karimi

February 18, 2020

They are mainly dealing with optimization and modeling of neural net:

1. Minibatch algorithm for training latent data models: MISSO algorithm. This is a paper I wrote with Eric and To but got rejected. I know it can be improved on two levels:
   - Empirically: New numerical experiments (bigger settings)
   - Theoretically: Simpler proof with milder assumptions. I believe you and maybe someone from the team familiar with optimization methods can have an impact on this too.
2. Following the SWA paper (https://arxiv.org/pdf/1803.05407.pdf) I want to investigate how averaging weights (or hyperparameters) of a Bayesian neural network (BNN) behave on the generalization error
   - Empirically: observing if SWA for BNN leads to better test error than SGD
   - Empirically: Checking if averaging is relevant only for BNN or if it can also work on non-neural net models (examples: pLSA, LDA etc). I also suspect that overparameterization is needed to observe such effect (this needs to be checked empirically first)
   - Theoretically: deriving PAC bounds on the generalization gap and show better results (averaging leading to better generalization has been proved in NIPS this year here: https://arxiv.org/pdf/1902.00744.pdf only for deterministic neural net).
3. On the same subject but a different study: understanding loss landscape of BNN. In https://arxiv.org/pdf/1806.06977.pdf, Gotmare et. al. investigate Mode connectivity for deterministic neural network, ie several independent NN trained in parallel are connected through a common low error curve. I would like to observe that for BNN. For BNNs the current main issue I see is robustness (i.e., it should be as easy or easier to get to work than normal deep learning models, but it is not at the moment), and efficiency (i.e., it should at most be 2-3x as slow as normal deep learning training, but currently it is slower) and those two topics above are I believe a step towards understanding what training a BNN means.

4. In https://arxiv.org/pdf/1906.07774.pdf Le Roux et. al. propose a method to understand generalization by looking at the loss landscape. They introduce the TIC criterion that measures in a nutshell the Hessian of the loss.

    - Empirically: I am thinking of deriving a cost efficient optimization algorithm using this TIC criteria to lead to better generalization
    - Theoretically: Using upper bounds on this criterion could help understanding generalization for overparameterized regime (d ¿¿ n).

5. Following my work on Variance reduced EM algorithms, I would like to investigate how Variance reduction can help the SAEM (a stochastic approximation version of the EM) reach better precision and/or converge faster. This algorithm is particularly well suited for latent data models and I use it a lot for Pharmacology data. The paper can be a general optimization paper (for NIPS) or specialized in Pharma modelling.

I will keep thinking of projects that can be done in the next few weeks to meet the NIPS deadline and will keep you updated. I am also open to any of your ideas or ongoing projects.