Understanding Synthetic Gradients and Decoupled Neural Interfaces ax1703.00522

2018/11/13 8:30 AM

- General analysis of the synthetic gradients methods – mostly empirical with some fake little theoretical proofs thrown about as well;
- testing seems to be mostly limited to linear SG modules
- main conclusion is that, when the target task is linearly separable, the linear SG module is able to approximate the true gradient quite well as shown by heat maps of the estimated loss function itself which can of course be derived by integrating the SG transformation function;
- this is mainly true when the MSE loss is used though not so much when the log loss is used because of the nonlinearity of the latter missed cannot be approximated well by linear SG;
- several existing alternatives to back propagation including feedback alignment, and direct feedback alignment and kickback have also been shown to be different forms of a more generalized version of SG where gradients from the SC module itself are also fed back into the main network;
- When using very deep heavily non-linear networks, using ST module in some cases seems to leads to convergence even faster than just using troop back propagation as long as the task is simple enough for example M NIST which is explained by the smoothing effect of approximating the loss by a quadratic function;
- this is also been shown to be a form of regularization which might prevent some overfitting when the network is too complex for the task