AM207 meeting summary

Notebook:	other		
Created:	11/1/2019 11:26 AM	**Updated:**	11/1/2019 12:25 PM
Author:	koko		

Meeting summary

Group member: Yuting Kou, Yiming Xu, Yizhou Wang, Ziyi Zhou
Project TF: Shu Xu
paper: Subspace Inference for Bayesian Deep Learning

Paper Content

1. summary

Bayesian neural network often involves high dimensional problem over the parameter space. In this paper, they designed a *Subspace Inference* which conduct dimension reduction over parameter space, then form posterior inference and Bayesian model average. This Subspace Inference shows great performance in terms of accuracy and likelihood and thus implies the low-dim sub-spaces contains a rich representations of full space.

2. Core: Bayesian Subspace Inference

1. Find a subspace

1. idea: find a subspace which contains high density and low-loss weight subspace.

$$W = \hat{W} + Pz$$

1. good starting point: pre-trained solution $\hat{W} = W_{SWA}$
2. methods:

1. Random Subspace: P is random vector from Normal Distribution

2. PCA over SGD trajectory: P is PCA components of weight deviation.
    1. SGD snapshot of trajectory: every c iterations
3. Curve Subspace:
    1. given good solution w_1, w_2, w_3, find low-loss curve path by

$$\min_{v_1,v_2} Loss_{nn}(w_i \text{ along the } \text{span}(v_1, v_2))$$

3. performance: Curve Subspace performs the best, but time consuming.

2. Conduct Bayesian Inference

1. goal: find posterior $p(z|D)$
2. multiple methods, we use HMC and BBB

3. Conduct Posterior Predictive

1. goal: evaluate how well the model is by reporting log likelihood and contour plots of posterior in subspace.

Potential interesting questions

1. Is there any dimension reduction method which performs in between PCA and Curve subspace?
2. Does initializations matters? How sensitive in initialization w_1, w_2, w_3 is curve subspace method?
3. Change SGD trajectory updates methods from `every c steps` to `low-loss`, what will the performance change?
4. Find connections of weight space to functional spaces.