SURROGATED ASSISTED BAYESIAN NEURAL NETWORK FOR GEOLOGICAL MODELS

Sean Luo

Supervisor: Professor Rohitash Chandra, Professor Richard Scalzo

School of Mathematics and Statistics
UNSW Sydney

January 2022

Submitted in partial fulfillment of the requirements of the degree of
Bachelor of Science with Honours

Plagiarism statement

I declare that this thesis is my own work, except where acknowledged, and has not been submitted for academic credit elsewhere.

I acknowledge that the assessor of this thesis may, for the purpose of assessing it:

- Reproduce it and provide a copy to another member of the University; and/or,
- Communicate a copy of it to a plagiarism checking service (which may then retain a copy of it on its database for the purpose of future plagiarism checking).

I certify that I have read and understood the University Rules in respect of Student Academic Misconduct, and am aware of any potential plagiarism penalties which may apply.

By signing this declaration I am agreeing to the statements and conditions above.

Signed: ____________________________ Date: ________________

Acknowledgements

By far the greatest thanks must go to my supervisor for the guidance, care and support they provided.

Thanks to my family for supporting me to advance in higher education.

Thanks go to Fred Flintstone and Robert Taggart for allowing his thesis style to be shamelessly copied.

Sean Luo, Day Month Year.

Abstract

This thesis is an investigation of the ?????????

Contents

Chapter 1
Introduction

In the past decade there has being some progression in the field of bayesian deep learning. Bayesian deep learning offers an intrinsic way to ensemble models that helps to quantify the posterior uncertainty in model parameters and prediction. Applying tradition bayesian inference techniques such as Monte Carlo sampling in deep neural networks have multiple challenges, mainly revolving around computational expenses due to the evaluation speed of large dataset, exponentially increasing rejection rate in MCMC sampler as the parameter space dimensions grows with large model, difficulty to evaluate uncertainty in realtime data, efficient proposal convergence and others. Variational bayesian inference methods are faster than sampling, but it does not offer an exact approximation of the target distribution.
Ideas explored:

Pure Bayesian:

1. Using stochastic mini batches by welling 2011
2. Using langevin dynamic gradient as proposal distribution, by chandra
3. Using parallel tempering to speed up exploration and exploitation by chandra [1]
4. Using surrogate models to simulate the posterior likelihood $p(D|\theta)$ to speed up evaluation, by chandra. [2]
5. Surrogate models are selected from variantional inference category, including, GAN, VAE, and Normaling Flow.
6. Modification to frequentist models that turns into bayesian:
7. Using dropout as intrinsic ensemble (bayesian?).
8. Using stochastic weight averaging withe gaussian noise (SWAG) - Wilson et al

This paper proposes to have the surrogate to generate weight and its corresponding numerator and denominator during accept reject to avoid all the extra forward pass and backward loss compare to normal DNN training. The input would be a window of past model parameters and a random vector from normal distrbution, The output would be a new set of model parameters, the corresponding numerator and denominator of accept reject step.

1.1 Generative Models

VAE is a shrinkage training model that has two shrinker that trains to a latent space of mean and variance of normal distribution. from the two latent vector it then trains to simulate a proper input.

GAN have discriminator and generator, generator samples from standard normal and maps it via a network to output space, discriminator compare the overall generated result to dataset. The training is conducted via a maximizing loss for discriminator and minimizing loss on generator.

Normalizing Flow is composed of multiple inversable layers that helps to provide a tractable posterior likelihood on weights. The target loss function for continuous case (we care for continuous case only as we plan to use it to generate the weights) is $L(D) \approxeq \frac{1}{N}\sum_{i=1}^{N} -\log(p_\theta(\tilde{x}^{(i)})) + c$. Where $\tilde{x}^{(i)} = x^i + u$ with $u \sim U(0, a)$ and $c = -M \times \log a$, a is dependent on data discretization and M is the dimension of x. This loss function encourages the network to learn to distribute data across the domain of the standard normal distribution. Specifically the GLOW model proposed by OpenAI [3] uses 1x1 convolution to boost training performance.

1.2 Notes:

Swag : approximates local loss distribution (posterior?) using standard normal momentum distribution. Propose -¿ Accept Reject: replacing forward pass for proposal Surrogate by Chandra et al: evaluate proposal and spit out its posterior likelihood i.e.

$min(1, pi(w*|x)/pi(wi|x) * q(wi|w*)/q(w*|wi))$

$min(1, p(x|w*)/p(x|wi) * p(w*|x)/p(wi|x) * q(wi|w*)/q(w*|wi))$

$p(w|x)$ and $q(wi|w*)$ are both easy to compute, the first is a loss, the second is proposal distribution (using tractable distribution like normal that s easy to evaluate) $p(x|w)$ requires forward pass, so we evaluate it using surrogate

Packages of interest include pymc, pyro, pytorch, (maybe sydney-machine-learning/pingala?)

Chapter 2

Sampling

Chapter 3
Method

Chapter 4

Experiment

Chapter 5

Analysis

Chapter 6

Discussion

Chapter 7

Conclusion

LaTeX [5] is a set of macros built atop TeX [4].

References

[1] Rohitash Chandra, Konark Jain, Ratneel Vikash Deo, and Sally Cripps. Langevin-gradient parallel tempering for bayesian neural learning. *CoRR*, abs/1811.04343, 2018.

[2] Rohitash Chandra, Konark Jain, and Arpit Kapoor. Surrogate-assisted parallel tempering for bayesian neural learning. *CoRR*, abs/1811.08687, 2018.

[3] Diederik P. Kingma and Prafulla Dhariwal. Glow: Generative flow with invertible 1x1 convolutions, 2018.

[4] Donald E. Knuth. *The TEX Book*. Addison-Wesley Professional, 1986.

[5] Leslie Lamport. *LATEX: a Document Preparation System*. Addison Wesley, Massachusetts, 2 edition, 1994.