Studying Uncertainity in Bayesian Neural Networks

Sharan Yalburgi, Julyan Arbel

July 16, 2019

1 Abstract

Availability of uncertainity measure is the one of the major reasons why Bayesian Neural Networks are topics of interest. It is therefore essential to understand the behavior of these uncertainity measures with varying prior conditions, amount of data and modelling tasks.

2 Introduction

Contributions of this paper is as follows:

3 Previous Work

3.1 Study of Bayesian Neural Networks with Non-Smooth Functions Imaizumi and Fukumizu, 2018

3.2 Consistency of Posterior Distributions for Neural Networks Lee, 2000

3.3 Classification kwith imperfect training labels Cannings, Fan, and Samworth, 2018

- Study effect of imperfect training labels on performance of classification methods.

- Bound the excess risk of an arbitrary classifier trained with imperfect labels in terms of its excess risk for predicting a noisy labe

3.4 A Theoretical Analysis of Deep Neural Networks for Texture Classification Basu et al., 2016

- Derives the upper bounds on the VC dimension of Convolutional Neural Network as well as Dropout and Dropconnect networks and the relation between excess error rate of Dropout and Dropconnect networks.

- Introduces concept of Intrinsic Dimension to show that texture datasets have a higher dimensionality than color/shape based data.

3.5 Understanding Priors in Bayesian Neural Networks at the Unit Level - Vladimirova et al., 2019

- Studies the "induced" prior distributions at a unit level.
    - I layer units are Gaussian
    - II layer units are sub-exponential
    - Units in deeper layers are characterized by sub-Weibull distributions.

- This paper is devoted to the investigation of hidden units prior distributions in Bayesian neural networks under the assumption of independent Gaussian weights.

3.6 Consistency of posterior distributions for neural networks - Lee, 2000

Considers only feedforward neural networks with a single hidden layer of units with logistic activation functions(not done for popularly used ReLU activations) and linear output unit(no Sigmoid at

the end). Shows that posterior probability of feed-forward neural networks is **"asymptotically consistent"**. $\hat{g_n}$ is asymptotically consistent for true regression function g_0 if

$$\int (\hat{g_n}(x) - g_0(x))^2 f_0(x) dx \xrightarrow{p} 0 \qquad (1)$$

p here means that the probability,

$P(|\int (\hat{g_n}(x) - g_0(x))^2 f_0(x) dx| > \epsilon) \rightarrow 0$ as $n \rightarrow \infty$.

Asymptotically consistence in terms of Hellinger neighbourhood: If (X_i, Y_i) f_0, the posterior is asymptotically consistence for f_0(true joint distribution over X & Y) over Hellinger neighbourhood if for every *epsilon* > 0

$$P(f : D_H(f, f_0) <= \epsilon | (X_i, Y_i) \forall i) \xrightarrow{p} 1 \qquad (2)$$

where Hellinger distance

$$D_H$$

is defined as

$$D_H(f, f_0) = \sqrt{\int \int (\sqrt{f(x,y)} - \sqrt{f_0(x,y)})^2 dx dy}$$

Another approach to consistency: Posterior probability of every neighbourhood of true function tends to 1, i.e, $P(A_\epsilon | Y) \longrightarrow 1$ as $n \rightarrow 1$.

Sieve Approach: The number of hidden nodes grows with the sample size so that asymptotically they use arbitrarily large number of nodes.

Parameter Approach: Number of hidden nodes taken as a parameter(not sure if as a hyperparameter) and show that joint posterior is also consistent. Extends earlier results on "universal approximation properties of neural networks to the Bayesian setting".

Shows mathematically that using a neural network to estimate and **continuous or square integrable** function will be consistent with probability tending to one given enough data. This is when the number of nodes grew with amount of data in a controlled way but $k(n) \rightarrow \infty$ as $n \rightarrow \infty$ where k is the number of nodes in the network or when the number of nodes is a parameter which can be estimated from the data.