Laboratories

SAND2019-7559C

Using Bayesian Neural Networks for UQ of Hyperspectral Image Target Detection

AUTHORS

Daniel Ries, Braden Smith,Josh Zollweg, Lauren Hund

Outline

1. Main points
2. What is a Bayesian neural network (BNN)?
3. BNN for target detection on hyperspectral image (HSI)
4. Areas of concern with BNN
5. Concluding remarks

Main Points

1. BNN give UQ providing a means to lower false alarm rates (among others)
2. There are currenty multiple problems with implementing BNN which directly affects the usefulness of (1)

Executive Summary: **BNN have potential in applications of high stakes prolems, but they are not ready for deployment**

What is a *Bayesian* Neural Network?

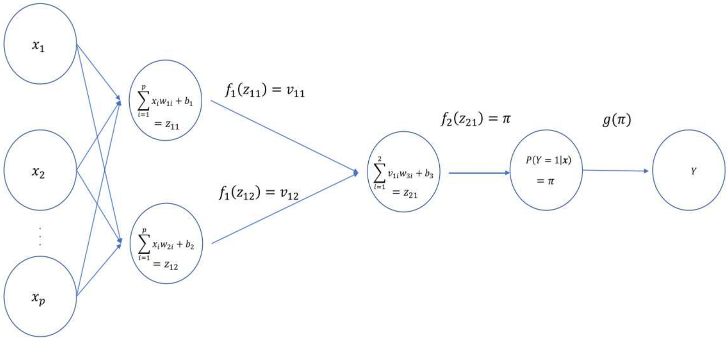
For those of you who "do machine learning, not statistics":
BNNs are simply NNs that have an implied loss function (you don't have to choose and you get UQ for free!)

For those of you who "do statistics, not machine learning":
BNNs are simply a highly nonlinear model where you put a prior on your parameters (and a Bernoulli likelihood for our binary data)

An Example BNN

$$Y_i \sim \text{Bernoulli}(\pi_i)$$
$$\pi_i = f_2(z_{21})$$
$$= f_2\left(\sum_{j=1}^{2} f_1\left(\sum_{k=1}^{p} x_k w_{jk} + b_j\right) w_{2j} + b_3\right)$$

$$w_{jk} \overset{iid}{\sim} N(0,1), \forall j, k$$
$$b_l \overset{iid}{\sim} N(0,1), l = 1, 2, 3$$



Estimating (or Training) BNNs

In traditional statistics, Bayesians sample their posterior distributions via MCMC
But this can be slowwww

Enter **Variational Inference**:

- Instead approximate posterior with $q(\theta)$ by solving:

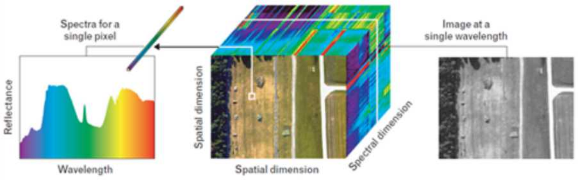
$$\underset{q^*}{\operatorname{argmin}}\, KL(q^*(\theta)||p(\mathbf{w},\mathbf{b}|y))$$

$$= \underset{q^*}{\operatorname{argmin}}\, KL(q^*(\theta)||p(\mathbf{w},\mathbf{b},y))$$

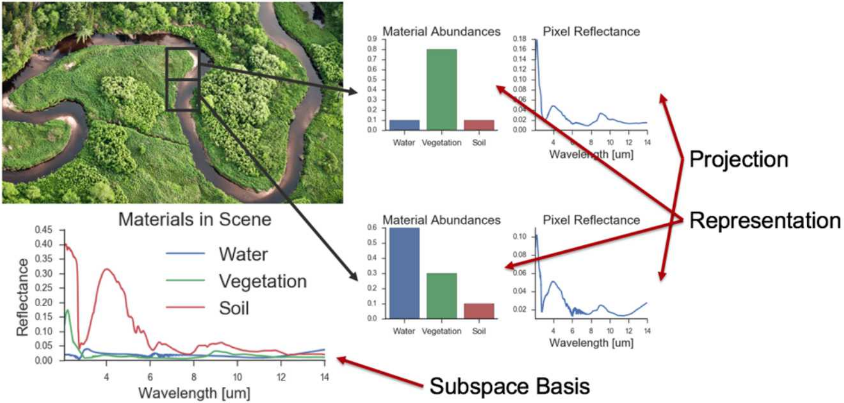
- Put restrictions on set of q^*
- Result is a loss function we can do gradient descent as usual on

Hyperspectral Images (HSI)

- Airborne spectrometers construct a (x, y, z) tensor
  - x and y describe the spatial dimension
  - z describes the spectrum at a single pixel (x, y)
- Specific materials are identified by their reflectance spectrum
- The target object might be smaller than the projection of a pixel on the ground



Mixing in HSI

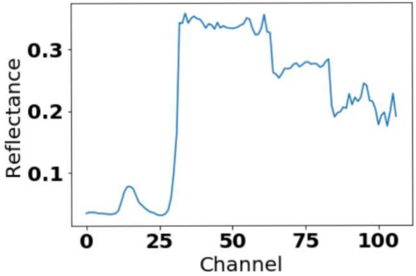


Megascene

- Simulated scene using DIRSIG model
- Green discs inserted as targets
  - 3 environments, and 3 times of day each
  - 125 targets in each scene
  - Targets range from large to subpixel (radius of 0.1m to 4m)
  - Some targets only partially visible

FPCA on Input

- Pixel level input is reflectance measured at 107 channels
- Use functional PCA to treat as functional
- 25 PCs were used



BNN on Megascene

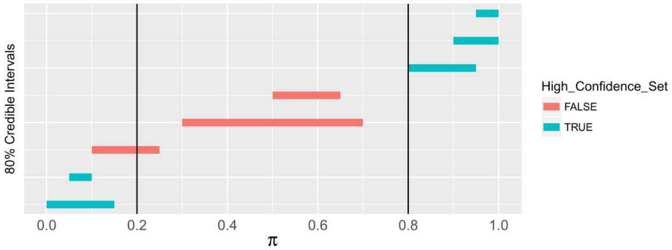
- **Architecture**: Input-10-5-2-Output
- **Activations**: Hyperbolic tan between hidden layers, inverse logit for output
- **Priors**: Standard normal on all weights and biases
- **Variational Distributions**: Independent normal

Leveraging UQ When Assessing Performance

We will evaluate test set performance in two cases:

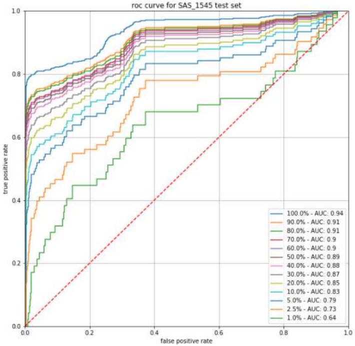
1. Using the full test set
2. Using a “high confidence” set

The “high confidence” set contains all pixels such that their 80% credible interval for $\pi \in (0, 0.2) \cup (0.8, 1)$

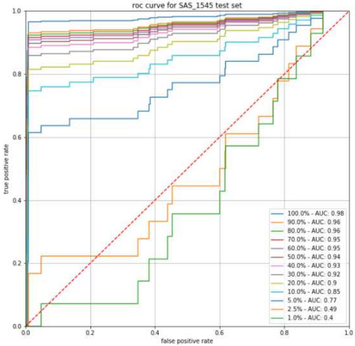


Test Set Performance

Full test set.



High confidence set.



Test Set Performance-Constant False Alarm Rate

A: Abundance of target in pixel

	A< 0.25	$0.25 \leq$ A < 0.75	A ≥ 0.75
Full Set	0.471	0.964	0.992
High Confidence	0.726	0.994	1.0

Table: Probability of detection for FAR = 0.05

Main Points

1. BNN give UQ providing a mean to lower false alarm rates (among others)
2. There are currenty multiple problems with implementing BNN which directly affects the usefulness of (1)

Executive Summary: **BNN have potential in applications of high stakes prolems, but they are not ready for deployment**

High Stakes Application

Many of problems Sandia (and others) works on are high stakes in nature

Please Stop Explaining Black Box Models for High-Stakes Decisions

Cynthia Rudin
Duke University
cynthia@cs.duke.edu

Abstract

Black box machine learning models are currently being used for high stakes decision-making throughout society, causing problems throughout healthcare, criminal justice, and in other domains. People have hoped that creating methods for explaining these black box models will alleviate some of these problems, but trying to *explain* black box models, rather than creating models that are *interpretable* in the first place, is likely to perpetuate bad practices and can potentially cause catastrophic harm to society. There is a way forward – it is to design models that

Areas of Concern when Implementing BNN

1. Posterior approximation via VI
2. Starting values
3. Difficulty of training

Problem 1: VI

MCMC is generally superior to VI for posterior estimation, computation aside

Asymptotically, MCMC and Full Rank VI will give the same answer, the true posterior

- Problems I work on do not have big data by any standard
- Full Rank VI is not very fast

Problem 1: VI

Mean Field VI with Normal variational distributions is common

- Often much faster
- Relatively easy to implement
- No chain convergence checking
- Asymptotic convergence of *marginal* posterior distributions

THIS IS A UQ WORKSHOP, WE KNOW THAT IS NOT ENOUGH

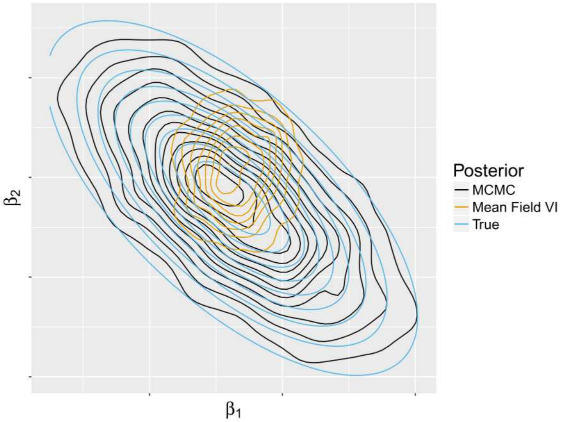
Problem 1: VI

What does “good” UQ mean?

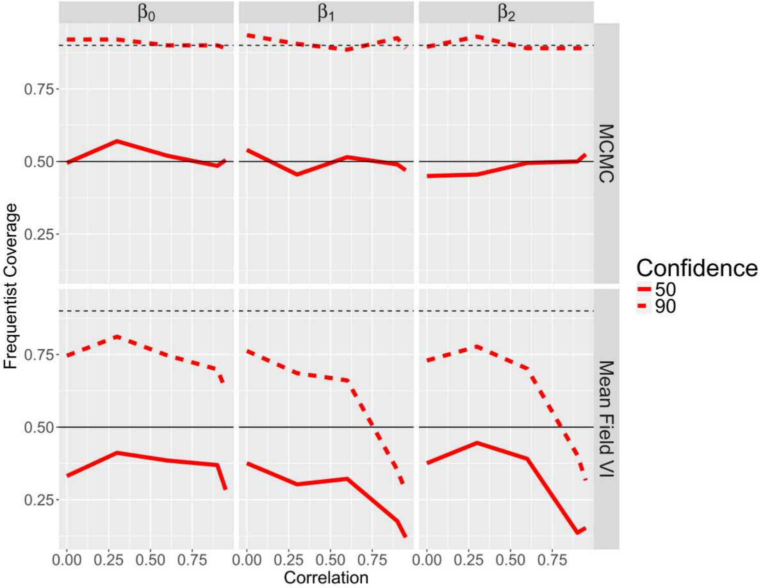
If you’re Bayesian, when your approximation to the posterior is *“good”*

Problem 1: VI

$$y_i|\beta, \sigma^2, x \sim N(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}, \sigma^2), i = 1, ..., 50$$
$$\sigma^2 \sim IG(a, b)$$
$$\beta|\sigma^2 \sim N(m, \sigma^2 V)$$



Problem 1: VI



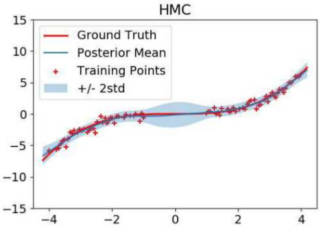
β_0
β_1
β_2
MCMC
Mean Field VI
Frequentist Coverage
0.25
0.50
0.75
Correlation
0.00
0.25
0.50
0.75
Confidence
50
90

Problem 1: VI

Another example of a bad posterior approximation...

Quality of Uncertainty Quantification for Bayesian Neural Network Inference

Jiayu Yao [*1] Weiwei Pan [*1] Soumya Ghosh [2] Finale Doshi-Velez [1]



Problem 2: Starting Values

There are well established ways to generate starting values for general NN, but not for BNN

Good Initializations of Variational Bayes for Deep Models

Simone Rossi [1] **Pietro Michiardi** [1] **Maurizio Filippone** [1]

Abstract

Stochastic variational inference is an established way to carry out approximate Bayesian inference for deep models. While there have been effective proposals for good initializations for loss minimization in deep learning, far less attention has been devoted to the issue of initialization of stochastic variational inference. We address this by proposing a novel layer-wise initialization strategy based on Bayesian linear models. The proposed method is extensively validated on regression and classification tasks, including

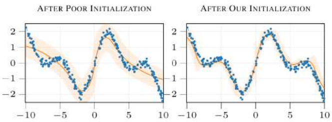


Figure 1: Due to poor initialization (**left**) SVI fails to converge even after 600+ epochs (RMSE = 0.613, MNLL = 29.4) while with our I-BLM (**right**) SVI easily recovers the function after few epochs (RMSE = 0.315, MNLL = −5.8). The architecture has three hidden layers with 500 neurons each, and uses the TANH activation function.

Problem 3: Difficulty of Training

BNN are known to be difficult to train

Possible reasons include:

- Poor estimation algorithm
- Poor starting values
- **Poor priors**

This results in using lower fidelity models

Conclusions

1. BNN provide a powerful UQ framework to approach high risk applications

2. BNN are not off-the-shelf ready for high risk applications

Thank you for listening!

dries@sandia.gov