

Get started

Open in app



ODSC - Open Data Science

28K Followers

About

Follow



Introduction to Bayesian Deep Learning



ODSC - Open Data Science Aug 5, 2020 · 6 min read

Table 1. Classification Error Rates on MNIST. * indicates result used an ensemble of 5 networks.

| Method | # Units/Layer | # Weights | Test Error |
|--|---------------|-----------|--------------|
| SGD, no regularisation (Sinz et al., 2003) | 800 | 1.3m | 1.6% |
| SGD, dropout (Hinton et al., 2012) | 800 | 1.3m | ≈ 1.3% |
| SGD, dropconnect (Wan et al., 2013) | 800 | 1.3m | 1.2%* |
| SGD | 400 | 500k | 1.83% |
| | 800 | 1.3m | 1.84% |
| | 1200 | 2.4m | 1.88% |
| SGD, dropout | 400 | 500k | 1.51% |
| | 800 | 1.3m | 1.33% |
| | 1200 | 2.4m | 1.36% |
| Bayes by Backprop, Gaussian | 400 | 500k | 1.82% |
| | 800 | 1.3m | 1.99% |
| | 1200 | 2.4m | 2.04% |
| Bayes by Backprop, Scale mixture | 400 | 500k | 1.36% |
| | 800 | 1.3m | 1.34% |
| | 1200 | 2.4m | 1.32% |

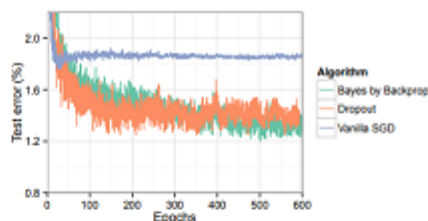
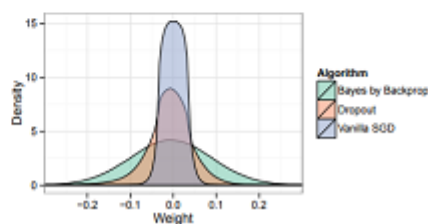


Figure 2. Test error on MNIST as training progresses.



Bayes' theorem is of fundamental importance to the field of data science, consisting of the disciplines: computer science, mathematical statistics, and probability. It is used to calculate the probability of an event occurring based on relevant existing information. **Bayesian inference** meanwhile leverages Bayes' theorem to update the probability of a hypothesis as additional data is encountered. But how can deep learning models benefit from Bayesian inference? A recent research paper written by New York University Assistant Professor Andrew Gordon Wilson addresses this question "[The Case for Bayesian Deep Learning](#)." As it turns out, supplementing deep learning with Bayesian thinking is a growth area of research.

[Get started](#)[Open in app](#)

of focus, current research efforts, and a look toward the future. It is common for Bayesian deep learning to essentially refer to Bayesian neural networks.

[Related article: [***Building Your First Bayesian Model in R***\]](#)

BDL Definitions

BDL is a discipline at the crossing between deep learning architectures and Bayesian probability theory. At the same time, Bayesian inference forms an important share of statistics and probabilistic machine learning (where probabilistic distributions are used to model the learning, uncertainty, and observable states).

The primary attraction of BDL is that it offers principled uncertainty estimates from deep learning architectures. Uncertainties in a neural network is a measure of how certain the model is with its prediction. With Bayesian modeling, there are two primary types of uncertainty:

- Aleatoric uncertainty — which measures the noise inherent in the observations, such as sensor noise which is uniform in the data set. This kind of uncertainty can't be reduced even with more data collected.
- Epistemic uncertainty — caused by the model itself, so it is also known as model uncertainty. It captures our lack of knowledge about which model generated our collected data. This kind of uncertainty can be reduced by collecting more data.



<https://odsc.com/>

[Get started](#)[Open in app](#)

probabilistic outputs. Epistemic uncertainty is modeled by placing a prior distribution over a model's weights and then capturing how much these weights vary over the data. On the other hand, Aleatoric uncertainty is modeled by placing a distribution over the outputs of the model.

Many data scientists believe that combining probabilistic machine learning, Bayesian learning, and neural networks represents a potentially beneficial practice, however, it's often difficult to train a Bayesian neural network. For training neural networks, the most popular approach is backpropagation, and for training BNNs, we typically use *Bayes by Backprop*. This method was introduced in the paper "[Weight Uncertainty in Neural Networks](#)," by Blundell, et al. for learning a probability distribution on the weights of a neural network. The following excerpt from the paper summarizes the approach:

"Instead of training a single network, the proposed method trains an ensemble of networks, where each network has its weights drawn from a shared, learnt probability distribution. Unlike other ensemble methods, our method typically only doubles the number of parameters yet trains an infinite ensemble using unbiased Monte Carlo estimates of the gradients."

In December 2019, there was a very compelling BDL [workshop](#) aligned with the NeurIPS 2019 conference. This site, with plenty of papers and slide presentations, represents a great learning resource for getting up to speed with BDL. The following is a published summary of the workshop that does a nice job of outlining the progress this field has been making:

"While deep learning has been revolutionary for machine learning, most modern deep learning models cannot represent their uncertainty nor take advantage of the well-studied tools of probability theory. This has started to change following recent developments of tools and techniques combining Bayesian approaches with deep learning. The intersection of the two fields has received great interest from the community, with the introduction of new deep learning models that take advantage of Bayesian techniques, and Bayesian models that incorporate deep learning elements. Many ideas from the 1990s are now being revisited in light of recent advances in the fields of approximate inference and deep learning, yielding many exciting new results."

[Get started](#)[Open in app](#)

Research into the area of BNNs dates back to 1990s with the following short-list of seminal papers in this burgeoning field:

- “[Keeping the neural networks simple by minimizing the description length of the weights](#),” by Geoffrey E. Hinton and Drew van Camp.
- “[Transforming Neural-Net Output Levels to Probability Distributions](#),” by John S. Denker and Yann leCun.
- “[Bayesian Learning for Neural Networks](#),” by Radford M. Neal
- “[A Practical Bayesian Framework for Backprop Networks](#),” by David J.C. MacKay.

Additionally, there is a growing [bibliography](#) available on research materials relating to BDL.

Areas of Focus for BNN Research

Understanding what a model does not know is a critical part of a machine learning application. Unfortunately, many deep learning algorithms in use today are typically unable to understand their uncertainty. The results of these models are often taken blindly and assumed to be accurate, which is not always the case.

It is clear to most data scientists that understanding uncertainty is important. So why isn't it done universally? The main issue is that traditional machine learning approaches to understanding uncertainty do not scale well for high dimensionality data like images and videos. To effectively understand this data, deep learning is needed, but deep learning struggles with model uncertainty. This is one reason for the rise in the appeal for BDL.

Another key property of BNNs is their connection with [deep ensembles](#). At a high level, both work to train a set of neural networks and yield predictions using some form of model averaging. One difference is that deep ensembles train these networks separately with different initializations while BNNs directly train a distribution of networks under Bayesian principles. Another difference is that deep ensembles directly average predictions from different networks while BNNs compute a weighted average using the

[Get started](#)[Open in app](#)

approximate Bayesian model average. Consequently, deep ensembles' success essentially brings both inspiration and additional insights to BNNs.

An additional area of focus relates to the BNNs use of probability distributions of weights instead of having deterministic weights. Due to a softmax function at the output layer to achieve the probability score, it reduces one class output probability score and maximizes the other. This leads to an overconfident decision for one class. This is one of the major difficulties with a point-estimate neural network.

Finally, Dropout is a widely-used regularization method that assists in reducing overfitting by randomly setting activations to zeros in a given layer. Dropout also can be used to make neural networks “Bayesian” in a straightforward manner, and in order to use it during inference, you just have to keep the Dropout, and sample several models (a process called MC dropout).

The Future Development of BNNs and BDL

[Related article: [From Idea to Insight: Using Bayesian Hierarchical Models to Predict Game Outcomes Part 1](#)]

The main hurdle for widespread adoption of BNNs and BDL in the past included computation efficiency and lack of publicly available packages. Recent encouraging development has taken a solid step over such hurdles. For example, there's been considerable work in terms of hardware and software to accelerate computation and new packages such as Edward have been specifically designed for probabilistic modeling and inference.

In the future, we can expect significant progress in BNNs for learning with small data, ensembles, along with model compression/pruning. In a broader sense, there also will be much more research based on the general nature of BDL, i.e. utilizing the reasoning ability of probabilistic graphical models for deep learning, in various problem domains such as computer vision, and natural language processing.

• • •

[Get started](#)[Open in app](#)

ODSC West 2020: “**The Bayesians are Coming! The Bayesians are Coming, to Time Series**” — This talk aims to allow people to update their own skill set in forecasting with these potentially Bayesian techniques.

ODSC Europe 2020: “**Bayesian Data Science: Probabilistic Programming**” — This tutorial will introduce the key concepts of probability distributions via hacker statistics, hands-on simulation, telling stories of the data-generation processes, Bayes’ rule, and Bayesian inference, all through hands-on coding and real-world examples.

[Data Science](#)[Artificial Intelligence](#)[Bayesian Statistics](#)[Deep Learning](#)[Machine Learning](#)[About](#) [Write](#) [Help](#) [Legal](#)

Get the Medium app

