Fast Computation of Uncertainty in Deep Learning

Mohammad Emtiyaz Khan
RIKEN Center for AI Project, Tokyo, Japan
http://emtiyaz.github.io

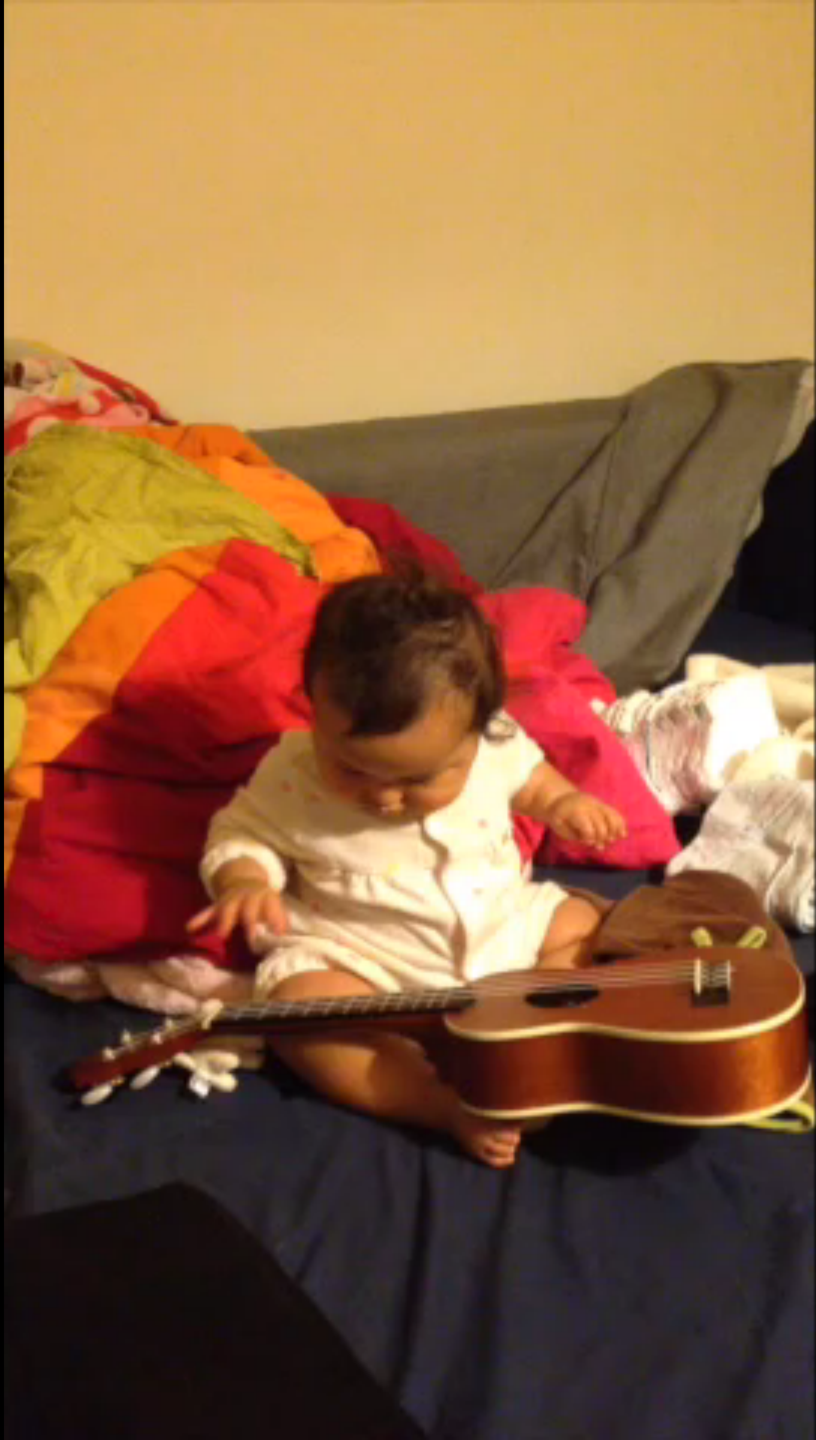
Joint work with
Wu Lin (UBC), Didrik Nielsen (RIKEN), Voot Tangkaratt (RIKEN)
Yarin Gal (University of Oxford), Akash Srivastava (University of Edinburgh)
Zuozhu Liu (SUTD, Singapore)

Long-term Goal

"To understand the fundamental principles of learning from data and use them to develop algorithms that can learn like living beings."

Learning by exploring

at the age of 6 months



Converged
at the age of
12 months

Transfer Learning at 14 months

Long-term Goal

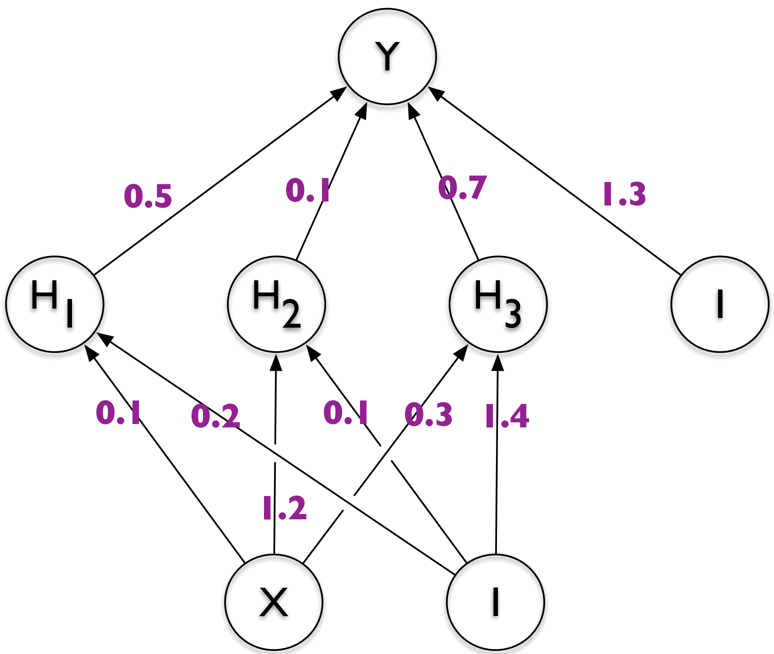
"To understand the fundamental principles of learning from data and use them to develop algorithms that can learn like living beings."

Outline

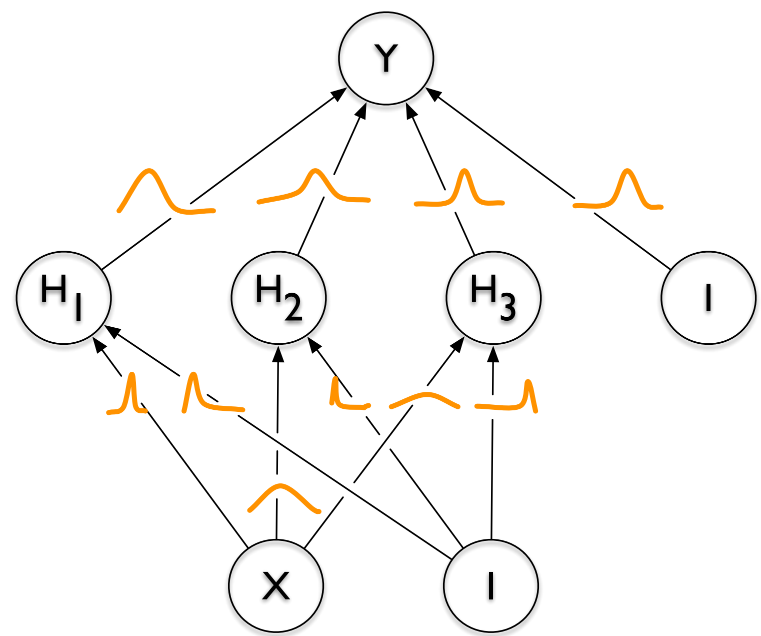
- Uncertainty
- “Bayesian” Deep Learning
- Fast computation of uncertainty
  - ICML 2018, Weight-perturbation in Adam to get uncertainty estimates (collaboration with University of Oxford and University of Edinburgh).
- Results

Uncertainty Estimation

Deep Learning



Bayesian Deep Learning

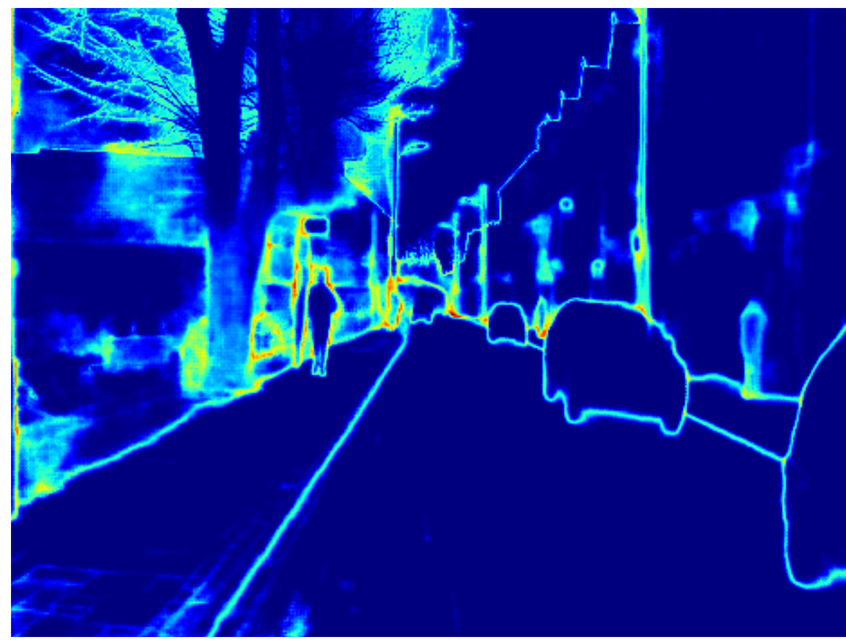


(taken from Blundell et al. 2015)

Example 1: Depth Estimation

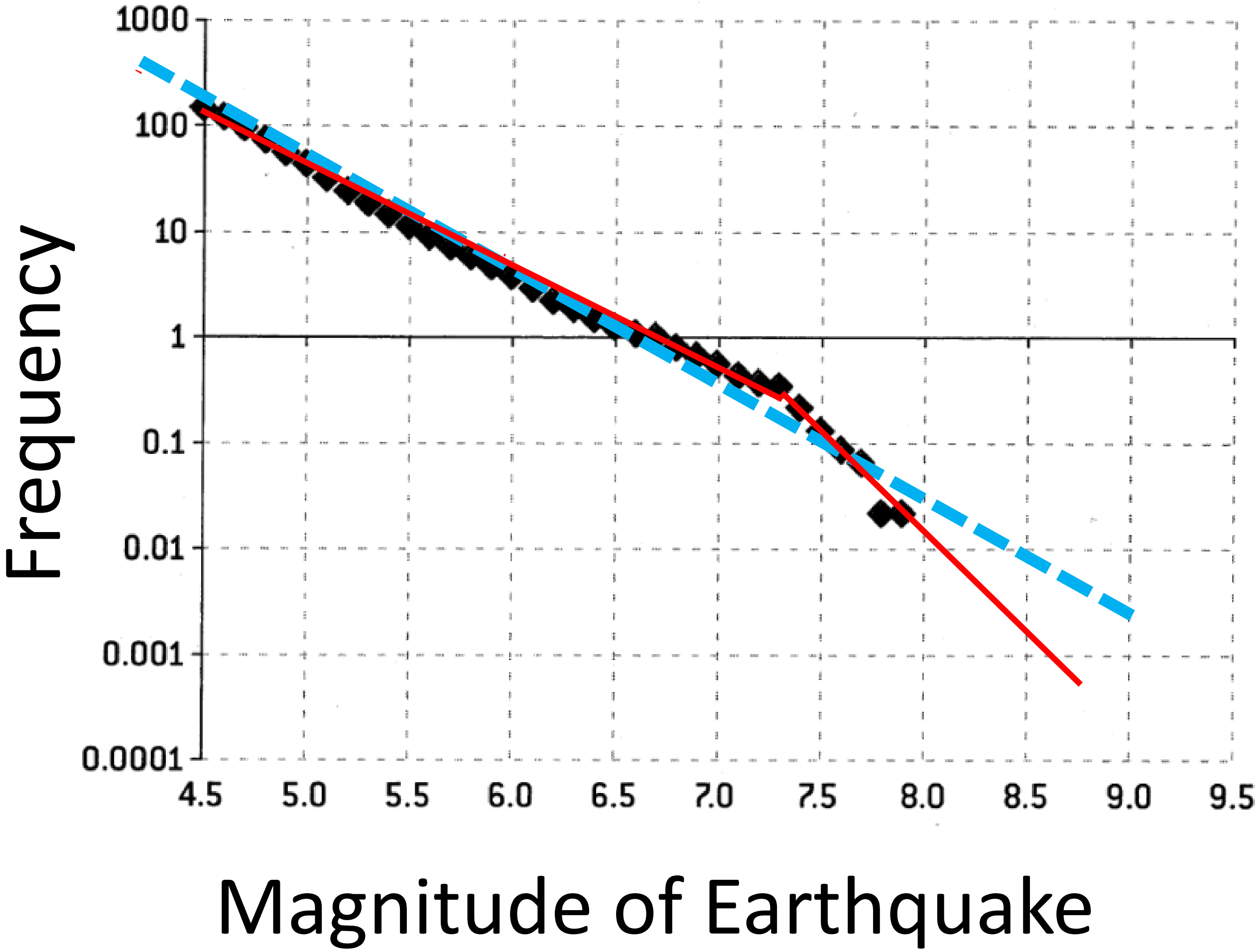
Scene

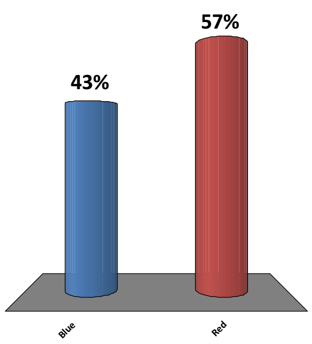
Uncertainty of depth estimates



(taken from Kendall et al. 2017)

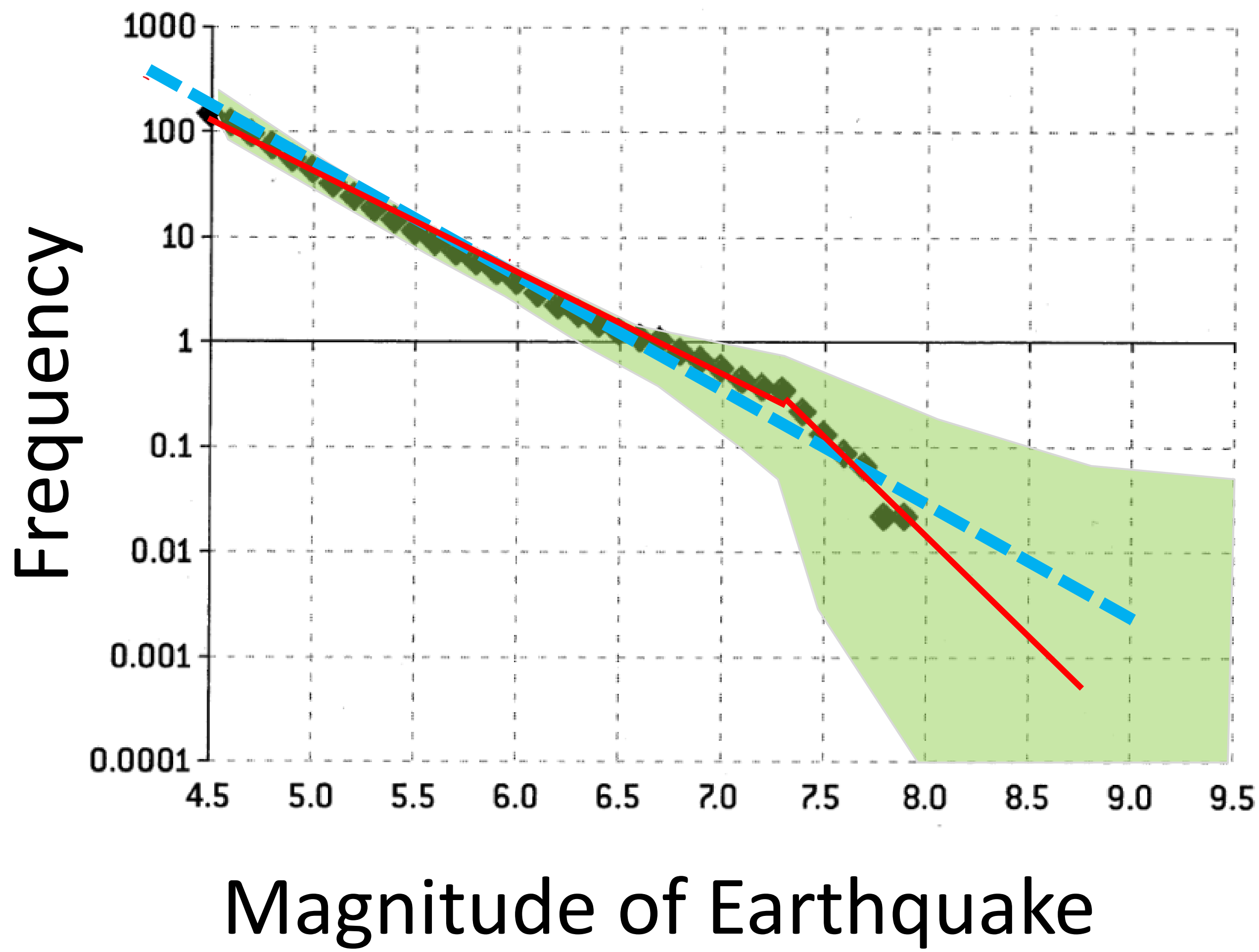
Example 2: Which is a Better Fit?





Real data from Tohoku (Japan). Example taken from Nate Silver's book "The signal and noise"

Example 2: Which is a Better Fit?



When the data is scarce and noisy, e.g., in medicine, and robotics.

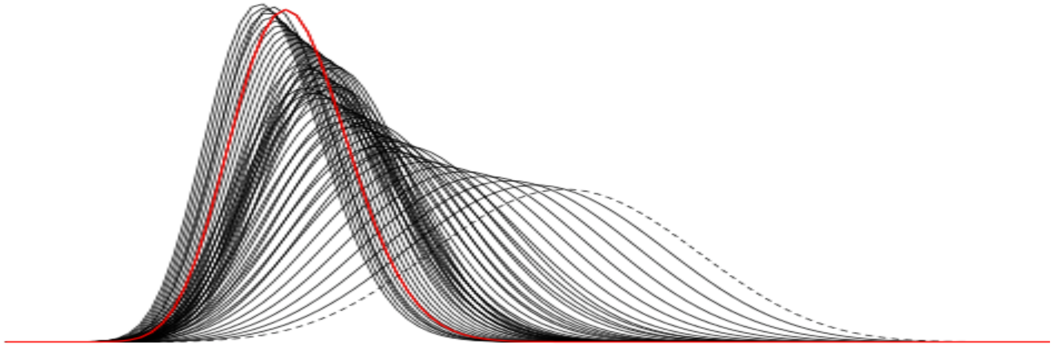
To Improve Deep Learning

Data-efficiency, robustness, active learning, continual/online learning, exploration

Bayesian Deep Learning

Compute a probability distribution over the unknowns given the data

“to know how much we don’t know”



Bayesian Inference is Difficult!

Data Parameters

Bayes' rule: $$p(w|\mathcal{D}) = \frac{p(\mathcal{D}|\theta)p(\theta)}{\underbrace{\int p(\mathcal{D}|\theta)p(\theta)dw}_{\text{Intractable integral}}}$$

- Approximate Bayesian Inference using gradient methods (SGD/Adam)
  - Gaussian VI: Bayes by Backprop (Blundell et al. 2015), Practical VI (Graves et al. 2011), Black-box VI (Rangnathan et al. 2014) and many more….
- Our work uses natural-gradient methods (faster and simpler than gradients methods)
  - Khan & Lin (AIstats 2017), Khan et al. (ICML 2018), Khan & Nielsen (ISITA2018)

Uncertainty Estimation with RMSprop

Model: $p(\mathcal{D}|\theta)\mathcal{N}(\theta|0, \lambda I)$

DNN Likelihood   Gaussian Prior

RMSprop for Deep Learning

$$\theta \leftarrow \mu$$
$$g \leftarrow \frac{1}{M}\sum_i \nabla_\theta \log p(\mathcal{D}_i|\theta)$$
$$s \leftarrow (1-\beta)s + \beta g^2$$
$$\mu \leftarrow \mu + \alpha \frac{g}{\sqrt{s}+\delta}$$

Variational RMSprop (Vprop) for Bayesian Deep Learning

$$\theta \leftarrow \mu + \epsilon, \text{ where } \epsilon \sim \mathcal{N}(0, Ns+\lambda)$$
$$g \leftarrow \frac{1}{M}\sum_i \nabla_\theta \log p(\mathcal{D}_i|\theta)$$
$$s \leftarrow (1-\beta)s + \beta g^2$$
$$\mu \leftarrow \mu + \alpha \frac{g+\lambda\mu/N}{\sqrt{s}+\lambda/N}$$

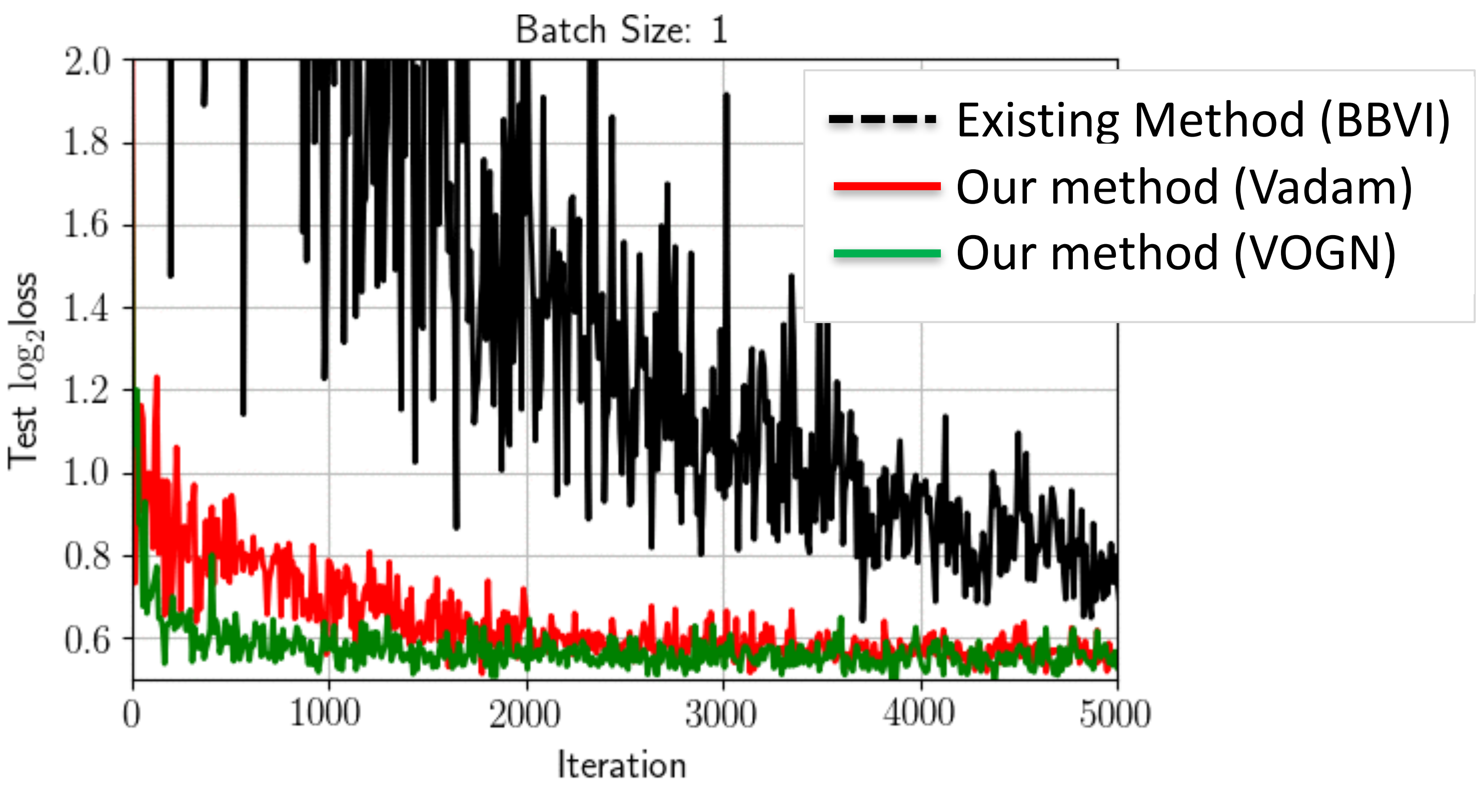
Note 1: Choose a small minibatch size.

Note 2: Similar version exist for Adam (Vadam)

Note 3: A better version is VOGN (details in the paper)

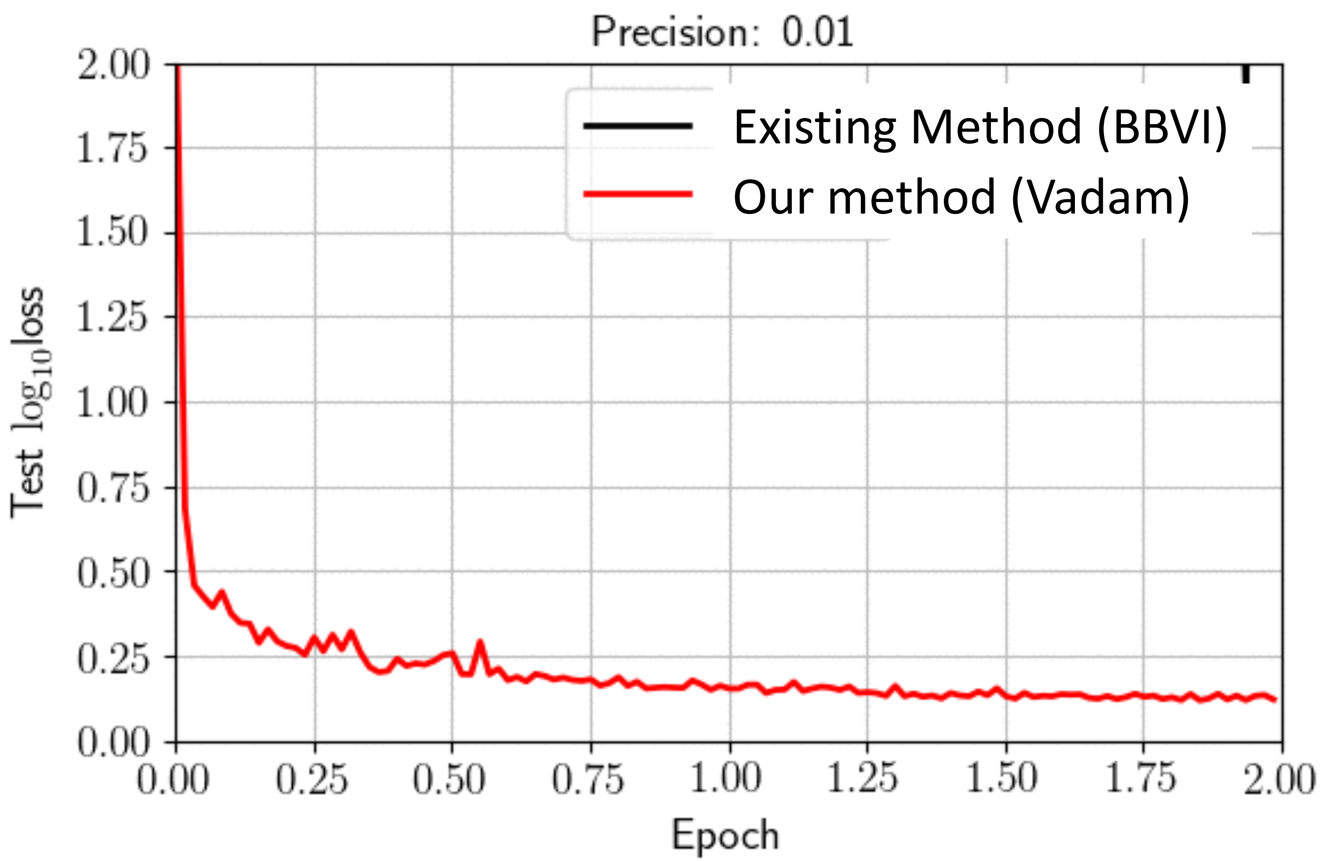
Faster, Simpler, and More Robust

Regression on Australian-Scale dataset using deep neural nets for various number of minibatch size.



Faster, Simpler, and More Robust

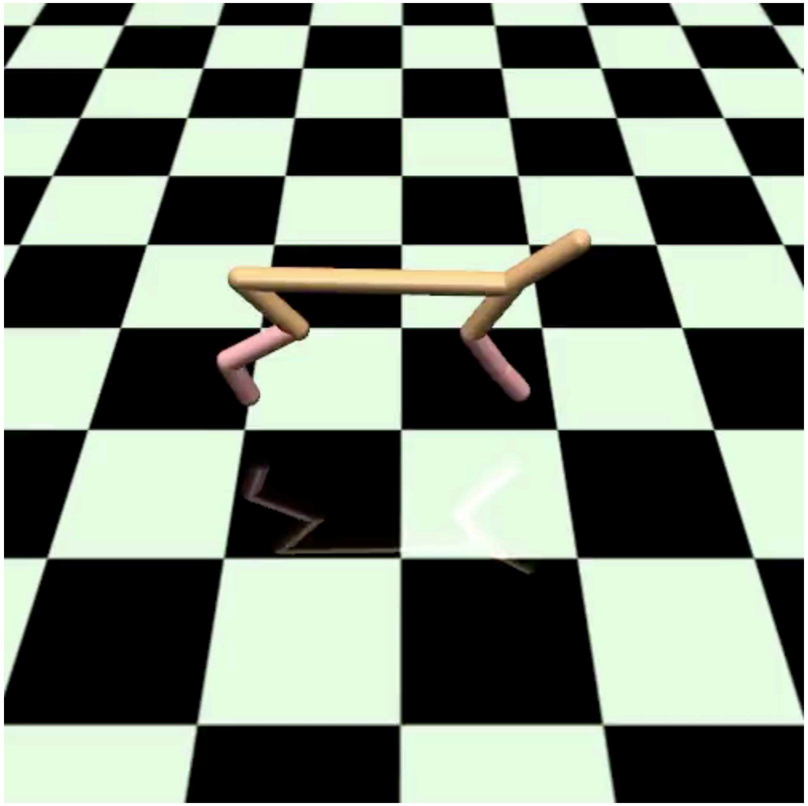
Results on MNIST digit classification (for various values of Gaussian prior precision parameter λ)



Deep Reinforcement Learning

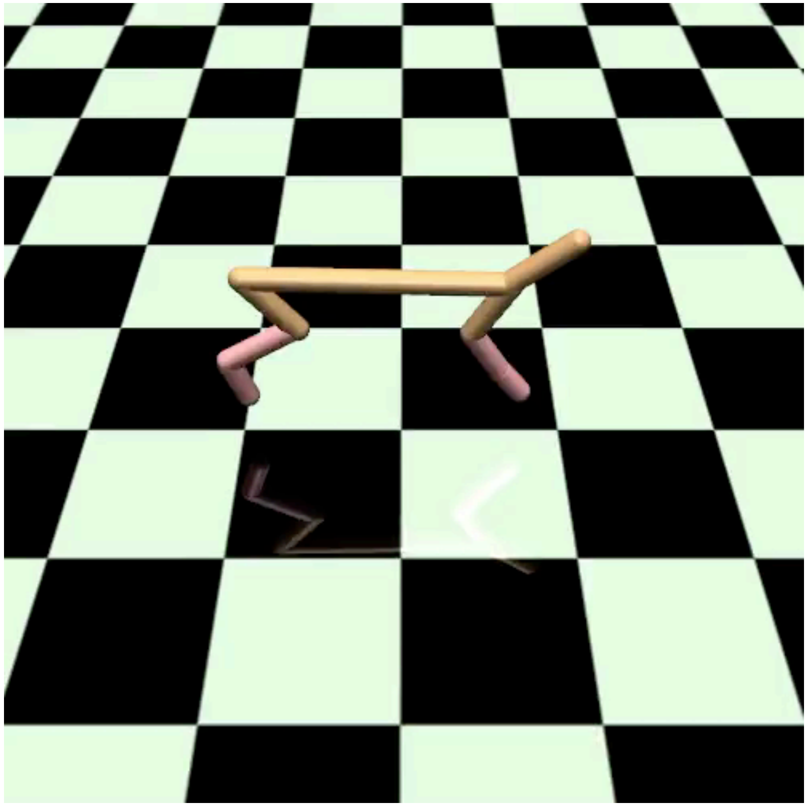
No Exploration (SGD)

Reward = 2860



Exploration using Vadam

Reward = 5264



Outline

- Uncertainty
- “Bayesian” Deep Learning
- Fast computation of uncertainty
  - ICML 2018, Weight-perturbation in Adam to get uncertainty estimates (collaboration with University of Oxford and University of Edinburgh).
- Results

References

https://emtiyaz.github.io

Fast yet Simple Natural-Gradient Descent for Variational Inference in Complex Models,
INVITED PAPER AT (ISITA 2018) **M.E. KHAN** and D. NIELSEN, [Pre-print]

Fast and Scalable Bayesian Deep Learning by Weight-Perturbation in Adam,
(ICML 2018) **M.E. KHAN**, D. NIELSEN, V. TANGKARATT, W. LIN, Y. GAL, AND A. SRIVASTAVA, [ArXiv Version] [Code]

Conjugate-Computation Variational Inference : Converting Variational Inference in Non-Conjugate Models to Inferences in Conjugate Models,
(AISTATS 2017) **M.E. KHAN** AND W. LIN [Paper] [Code for Logistic Reg + GPs] [Code for Correlated Topic Model]

Thanks!

https://emtiyaz.github.io