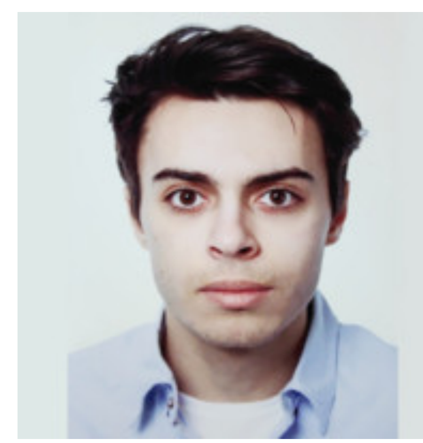
International Conference on Machine Learning (ICML 2021)

Bayesian Deep Learning via Subnetwork Inference

Erik Daxberger | Eric Nalisnick | James Allingham | Javier Antorán | Miguel Hernández-Lobato




UNIVERSITY OF CAMBRIDGE

Max Planck Institute for Intelligent Systems

UNIVERSITY OF AMSTERDAM

Motivation

Problem: Deep neural networks (DNNs) make **overconfident** predictions

Principled solution: Quantify predictive uncertainty via *Bayesian deep learning*

Motivation

Goal: Infer *posterior distribution* over DNN weights

Problem: Modern DNNs are too big!

OpenAI debuts gigantic GPT-3 language model with 175 billion parameters

Solution(?): Make strong assumptions, e.g. independence between weights

➡ Deteriorates quality of induced uncertainty estimates!
(Ovadia 2019, Fort 2019, Foong 2019, Ashukha 2020)

(source: https://venturebeat.com/2020/05/29/openai-debuts-gigantic-gpt-3-language-model-with-175-billion-parameters/)

Motivation

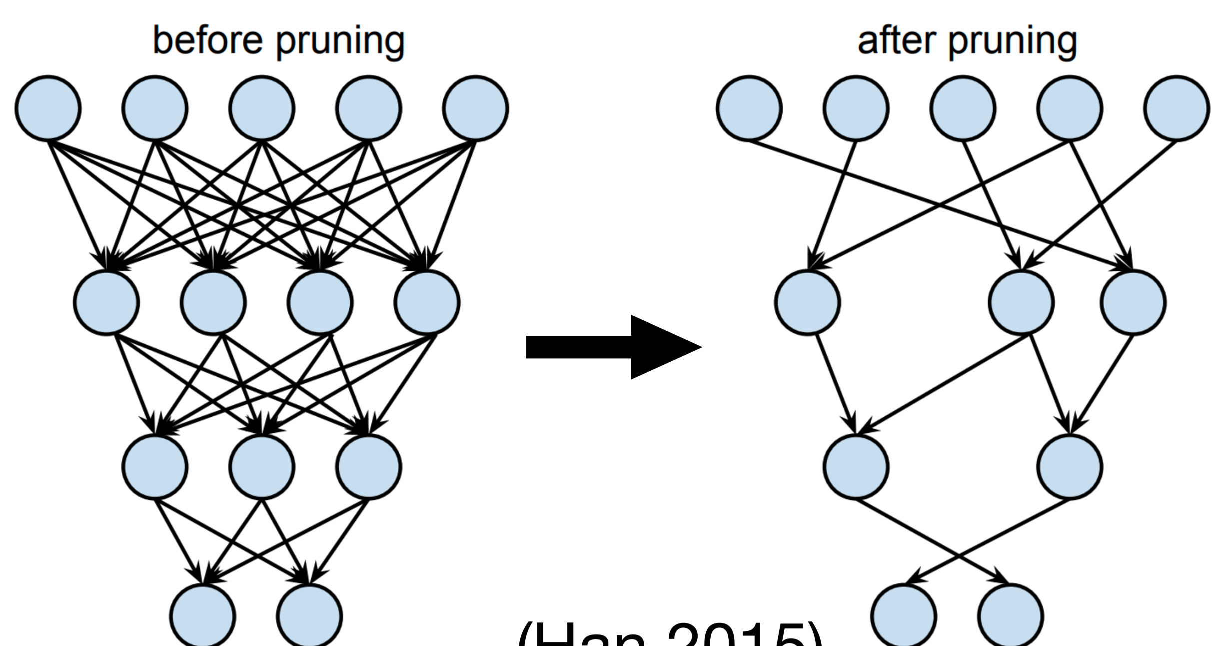
Observation: Almost all Bayesian deep learning methods try to do inference over **all** the weights of the DNN.

Do we really need to estimate a posterior over **ALL** the weights?!

Idea

Observation: Due to overparameterization, a DNNs **accuracy** is well-preserved by a **small subnetwork**

How to find those subnetworks? —> DNN **pruning**, e.g. (Frankle & Carbin 2019)



(Han 2015)

Question: **Can a full DNN's *model uncertainty* be well-preserved by a *small subnetwork's* model uncertainty?**

Answer: **Yes!**

Subnetwork Inference

1. Point Estimation

—> do standard SGD model training

2. Subnetwork Selection

—> minimize discrepancy between subnetwork posterior & full posterior

3. Bayesian Inference

—> infer full-covariance Gaussian posterior via Laplace approximation

4. Prediction

—> use all weights: integrate over the subnetwork & keep other weights fixed

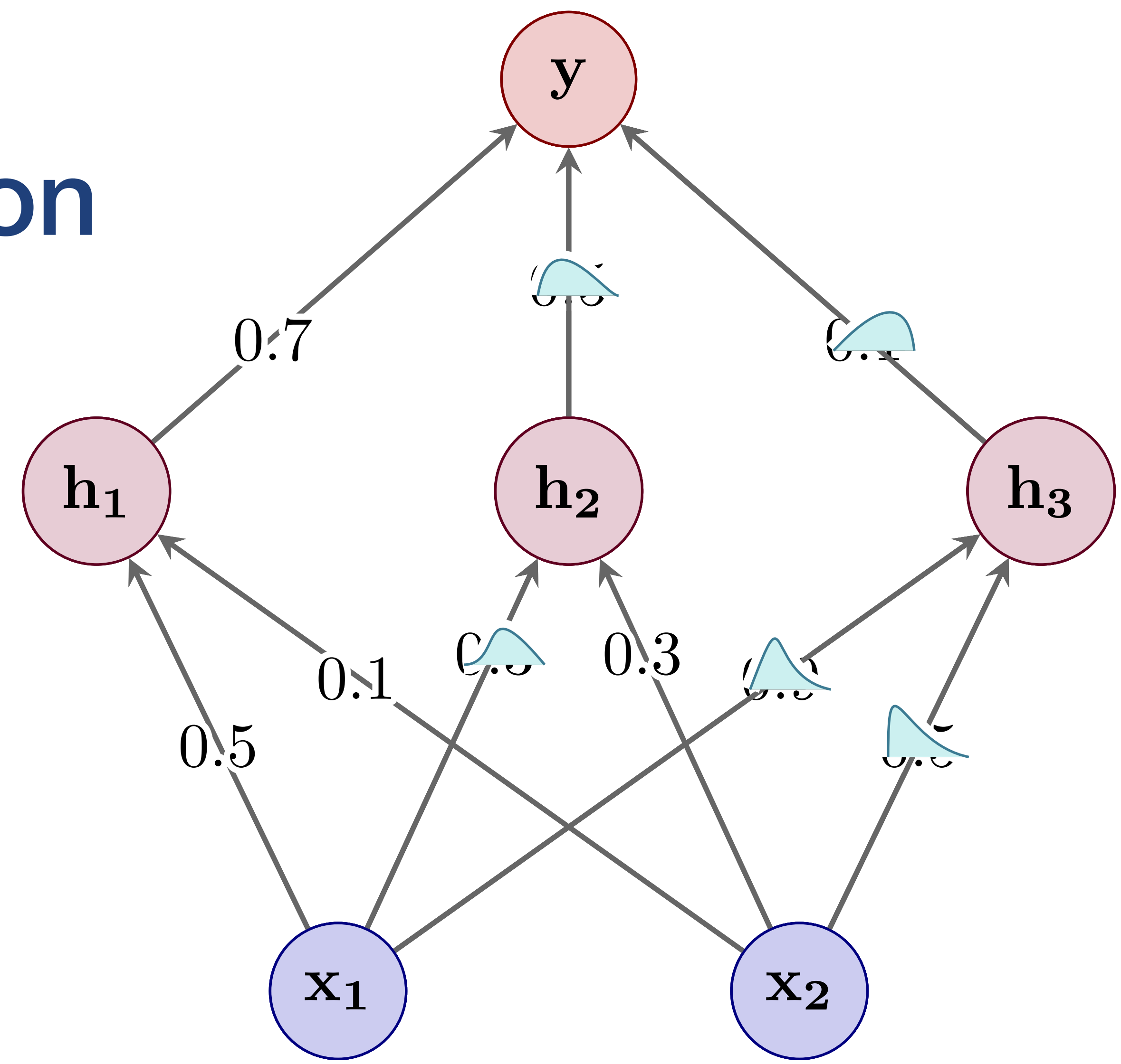


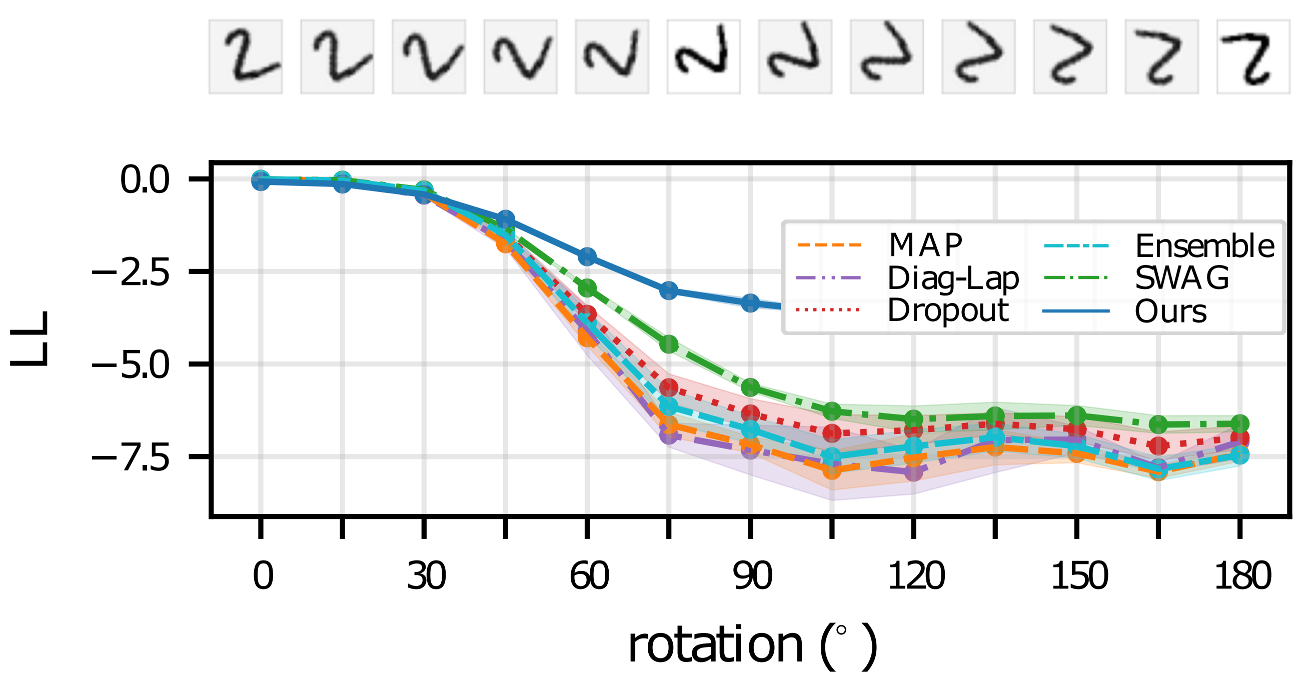
Image Class. under Distribution Shift

Model:
ResNet-18 with **11M** weights

Wasserstein subnetwork inference
subnet of just **40K (0.4%)** weights

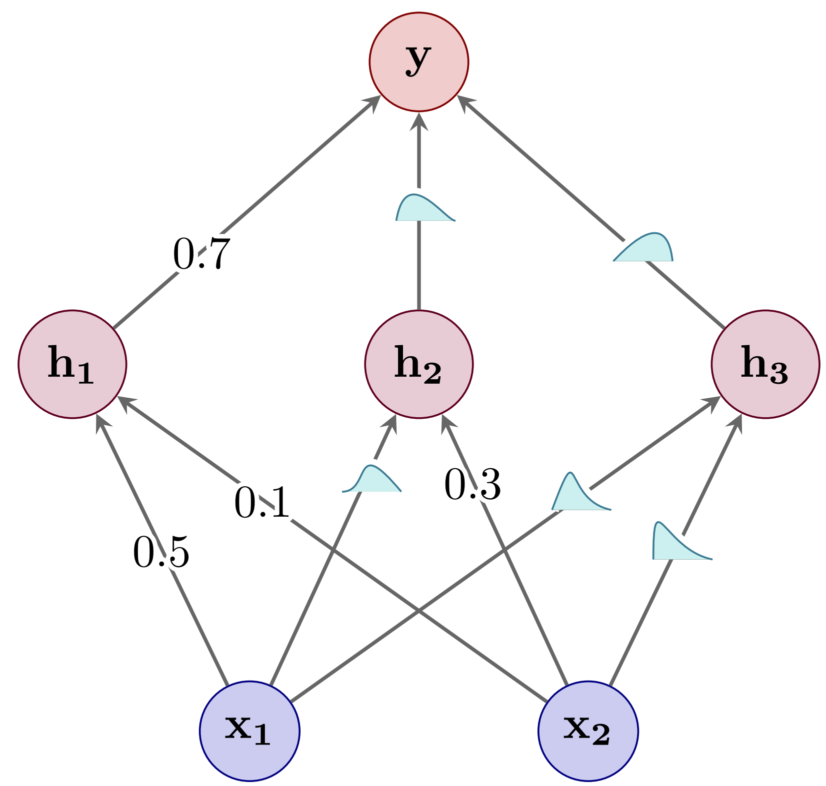
Baselines:
- MAP
- Diagonal Laplace
- MC Dropout (Gal 2016)
- Deep Ensembles (Lakshminarayanan 2017)
- SWAG (Maddox 2019)

Rotated MNIST (Ovadia 2019)



Subnet inference is **more robust to distribution shift** than popular baselines!

Take-Home Message



We propose a Bayesian deep learning method
that does ***expressive inference***
over a carefully chosen ***subnetwork***
within a neural network,
and show that this ***performs better*** than
doing crude inference over the full network.