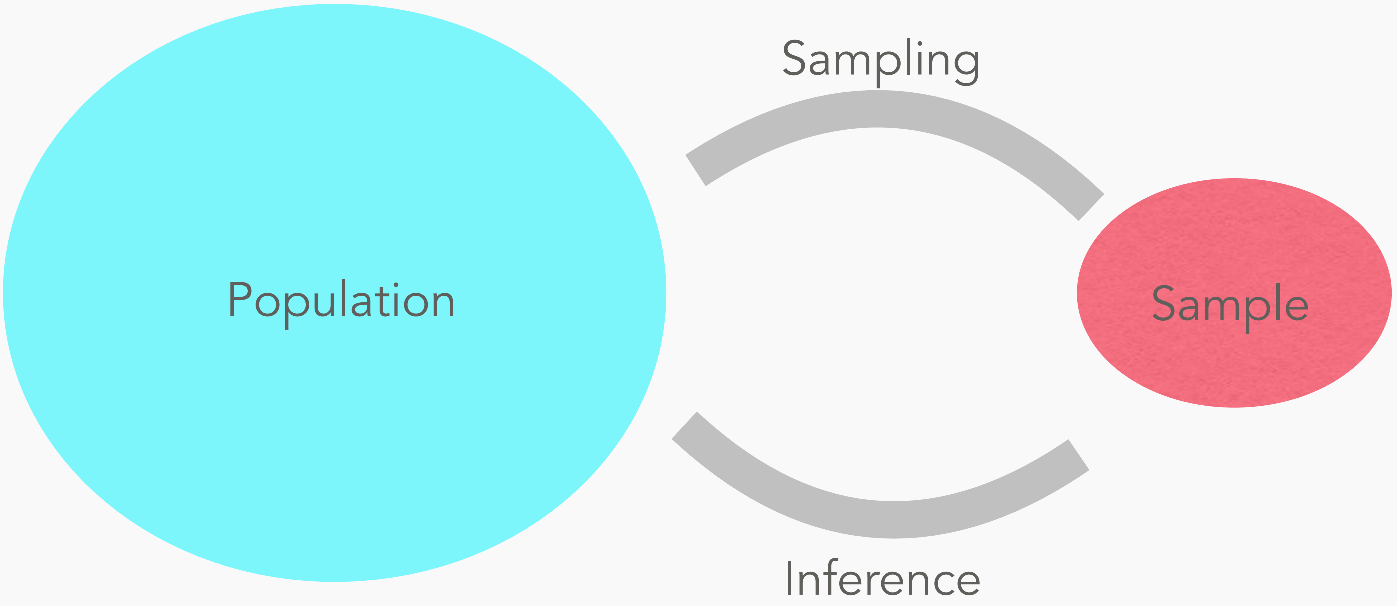
Advanced Section: Variational Inference

CS109B Data Science 2

Pavlos Protopapas, Mark Glickman and Chris Tanner

Statistical Inference

Draw conclusions about an underlying distribution of probabilities from a sample



Outline

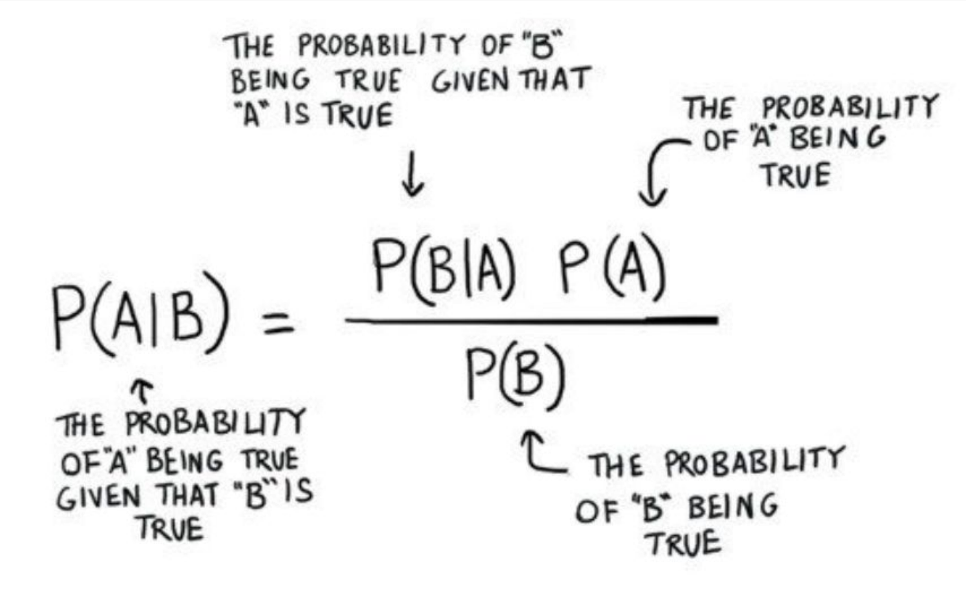
1. Bayesian Inference
2. Markov Chain Monte Carlo
3. Bayesian Neural Networks
4. Variational Inference
5. Drop Out as a Bayesian Approximation
6. Bootstrap for Inference

Bayesian Inference

Probability as a measure of *believability in an event*

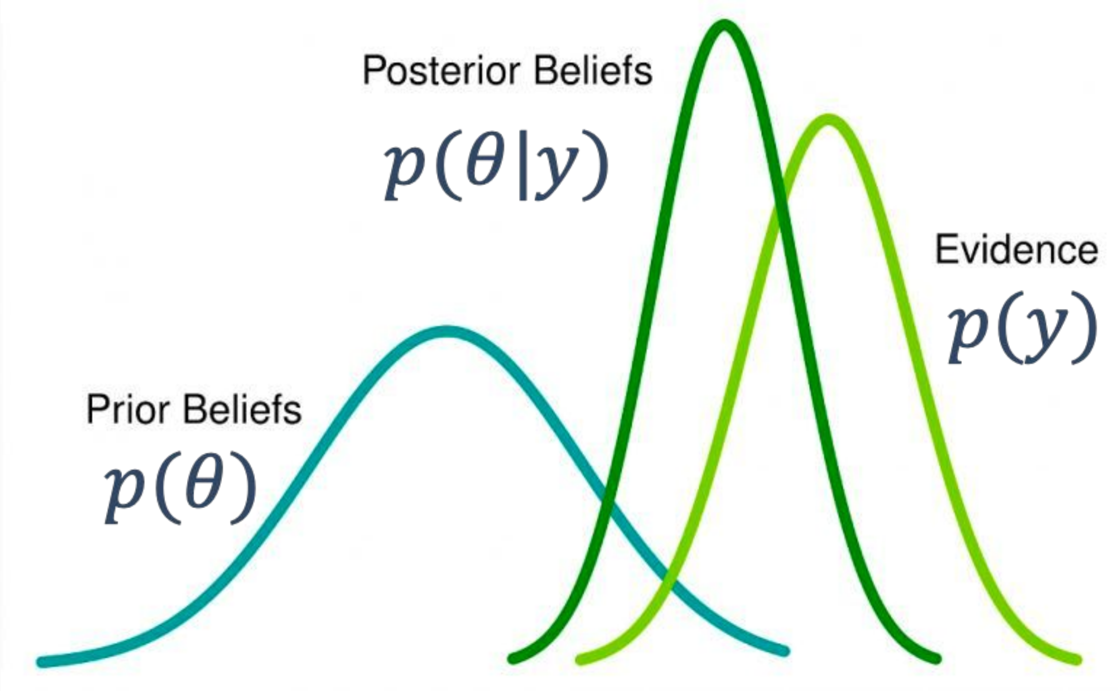
$$p(\theta|y) \propto p(y|\theta)p(\theta)$$

Model Data



Bayesian Inference

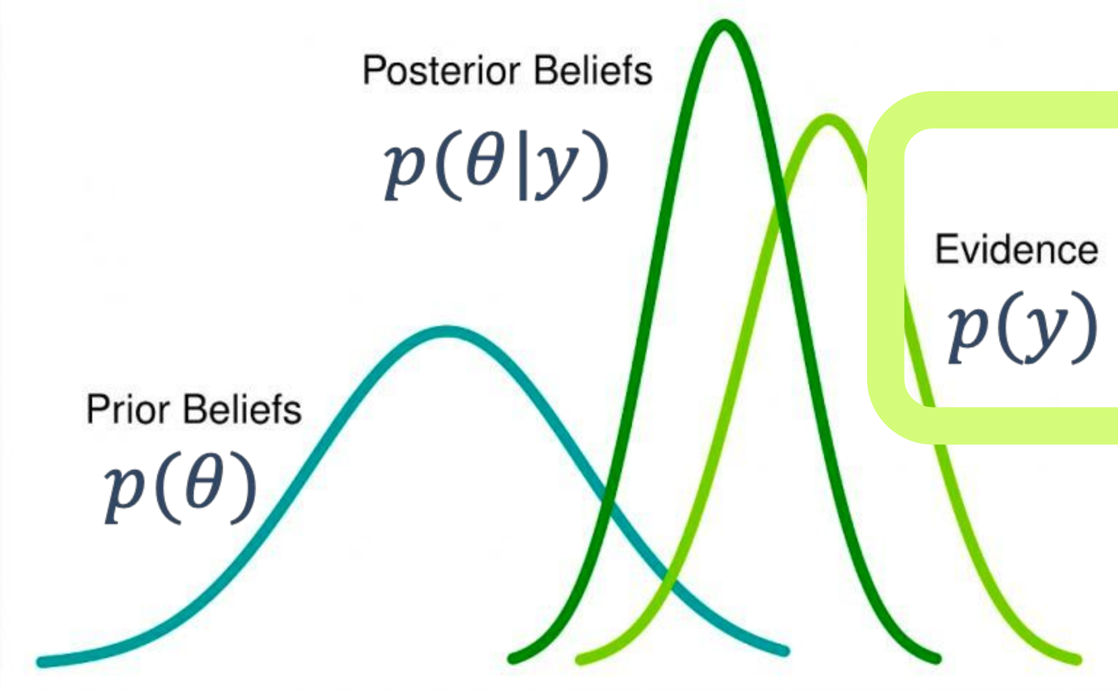
$$p(\theta|y) \propto p(y|\theta)p(\theta$$



“When the facts change, I change my mind. What do you do, sir? “John Maynard Keynes

Bayesian Inference

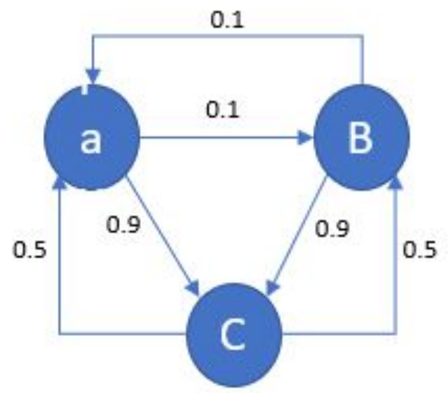
$\boxed{p(\theta|y) \propto p(y|\theta)p(\theta}$



Intractable

$p(y) = \int_\theta p(y|\theta)p$

"When the facts change, I change my mind. What do you do, sir? "John Maynard Keynes

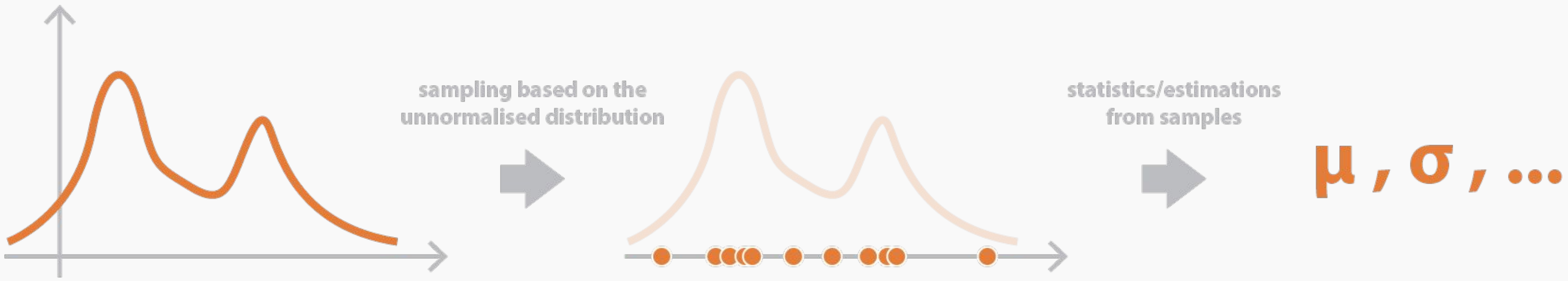


$$P = \begin{bmatrix} 0 & 0.1 & 0.9 \\ 0.1 & 0.0 & 0.9 \\ 0.5 & 0.5 & 0.0 \end{bmatrix}$$

$$p(z^{(m+1)}|z^{(1)},\ldots,z^{(m)}) = p(z^{(m+1)}|$$

$$p(z^{(m+1)}) = \sum_{z^{(m)}} p(z^{(m+1)}|z^{(m)})p$$

MCMC: Sampling method



Unnormalised distribution
whose normalisation factor
computation is intractable

Samples
that can be obtained with MCMC and
without proceeding to the normalisation

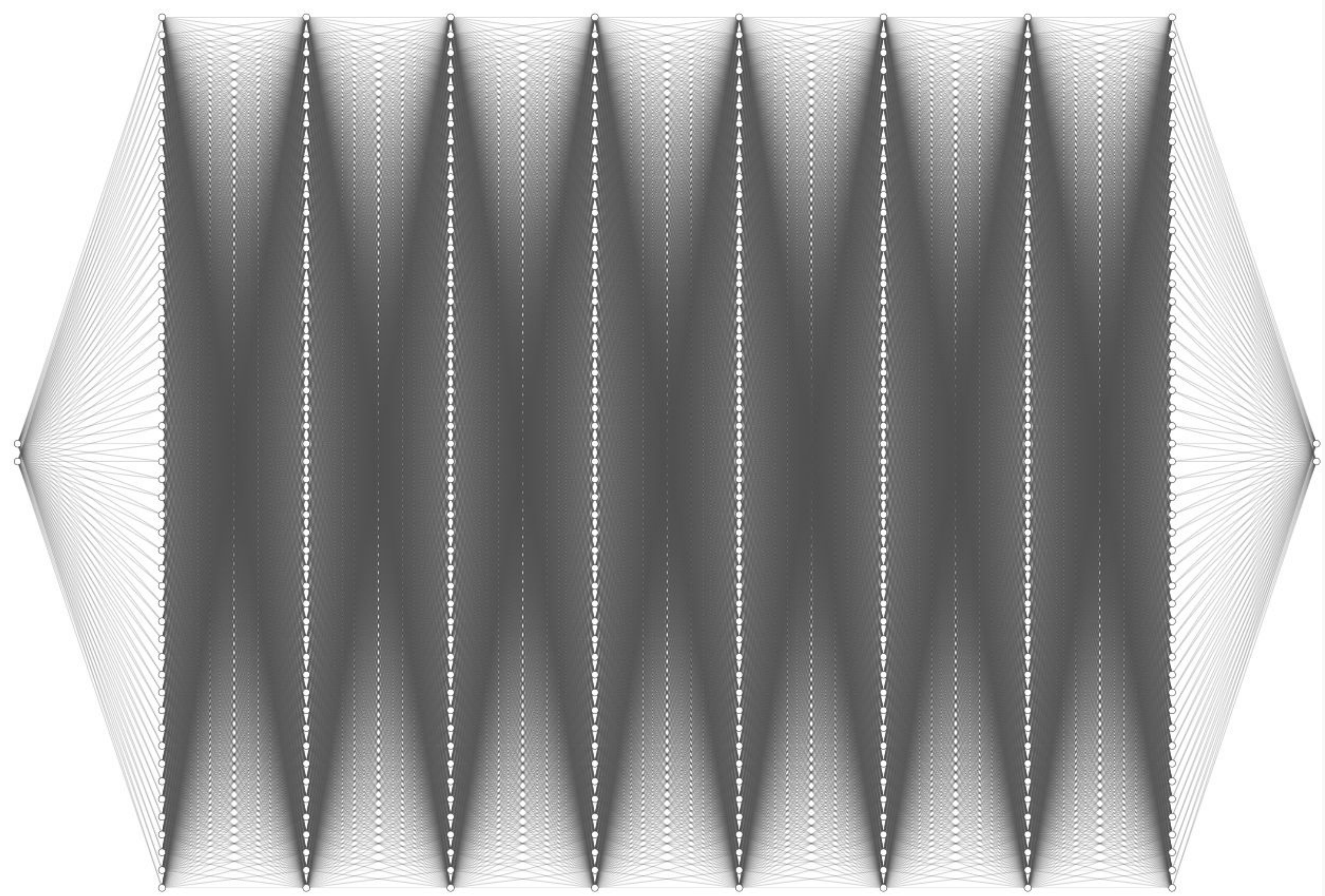
Statistics or estimations
that can be computed based on
the generated samples

MCMC

Credit: Towards Data Science

Bayesian Neural Networks: FCNN

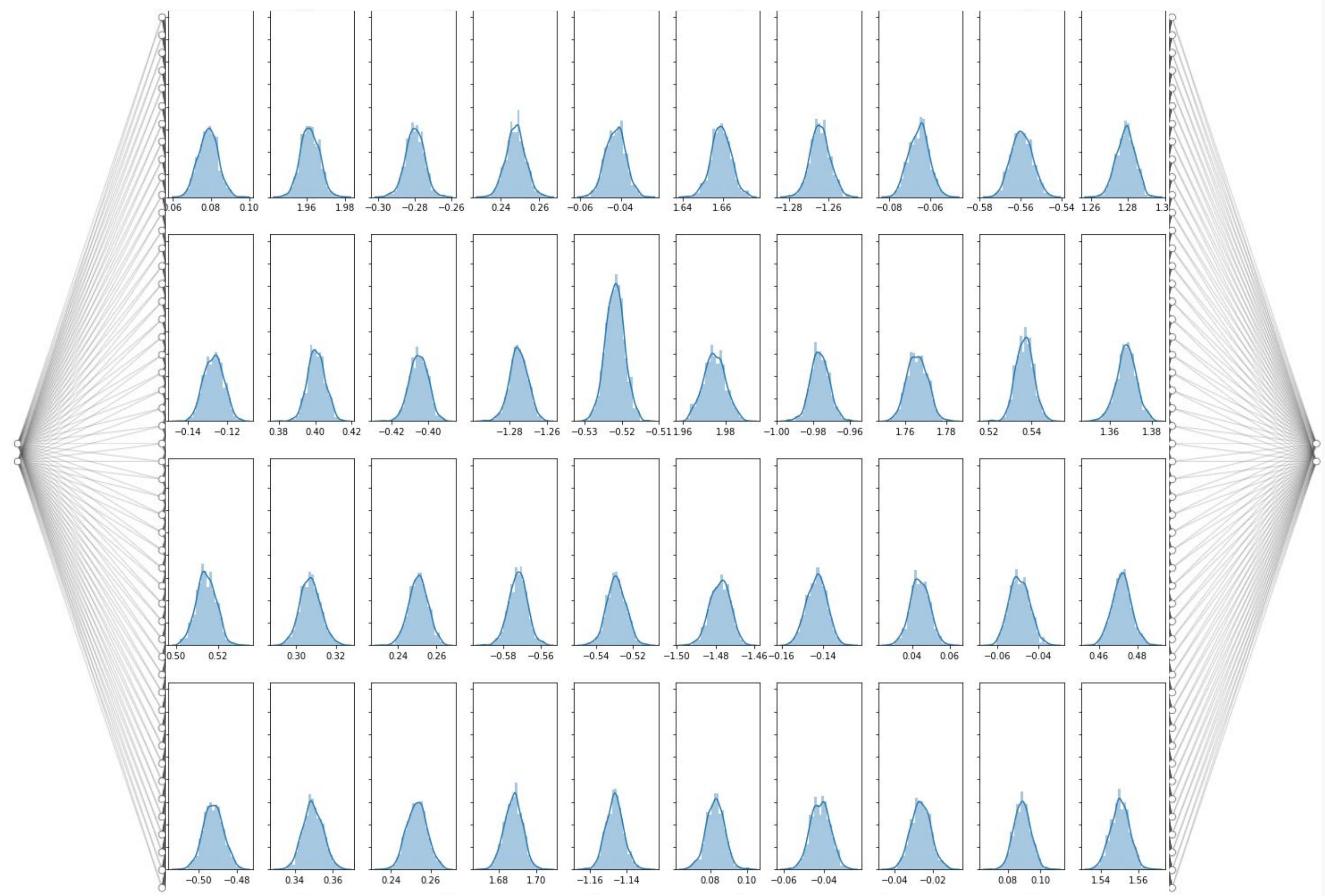
Input →



→ Output

Bayesian Neural Networks: FCNN

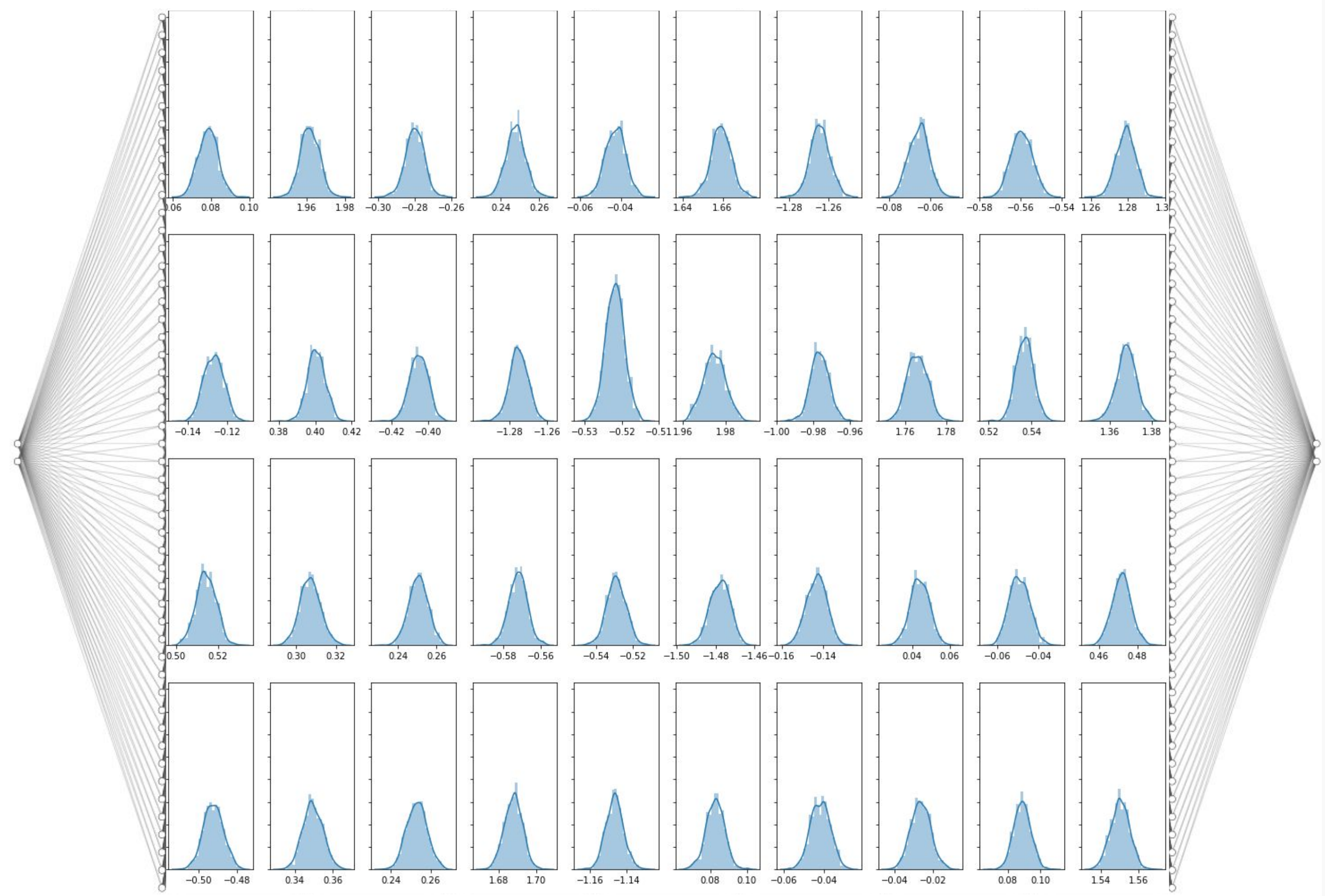
Input →



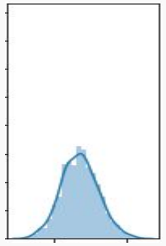
→ Output

Bayesian Neural Networks

Input



Output



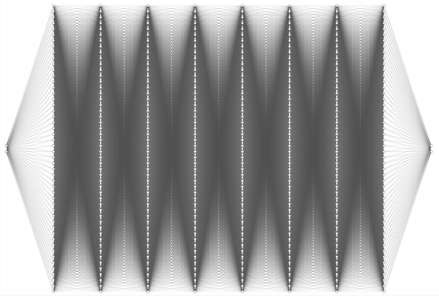
Bayesian Neural Networks

Priors

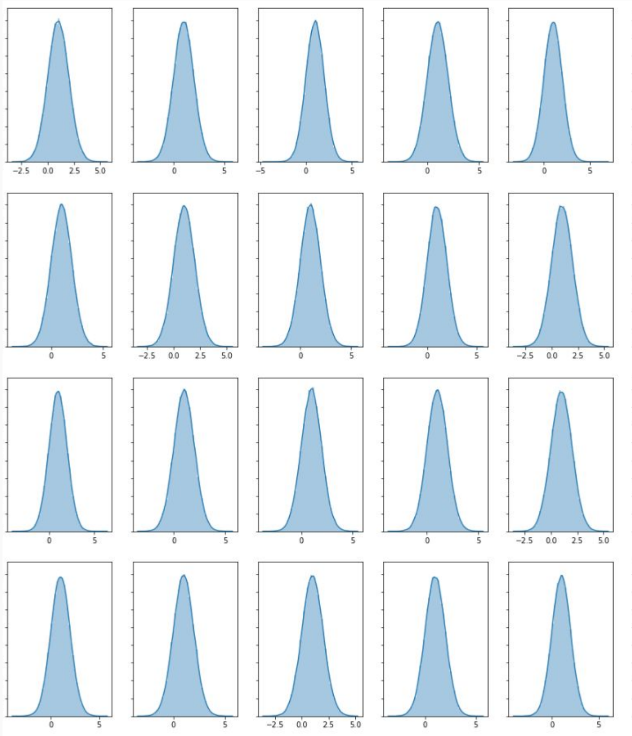
$p(\theta) \times p(y|\theta) \propto p(\theta|y)$

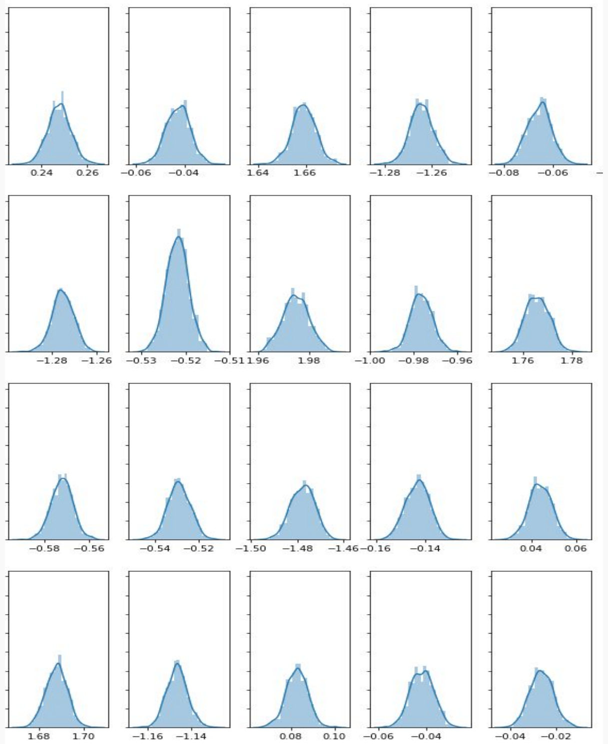
Means

FCNN



& Scale





Bayesian Neural Networks

$$p(\theta) \times p(y|\theta) \rightarrow p(\theta|y)$$

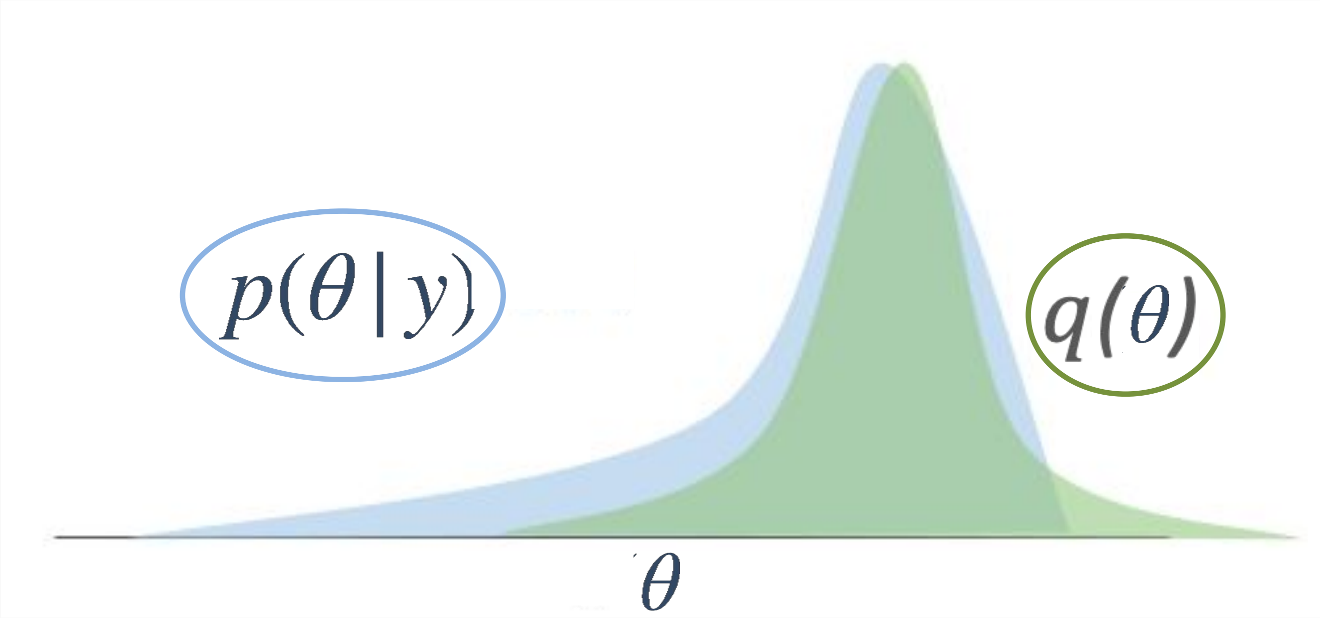
MCMC is eventually accurate, but not scalable to large models

Approximate Bayesian Inference: Variational Inference

Variational Inference

Optimization approach -> Q a family of “nice” distributions

$$p(\theta \mid y) = \frac{p(y \mid \theta)\, p(\theta)}{\int p(y, \theta)\, d\theta}$$



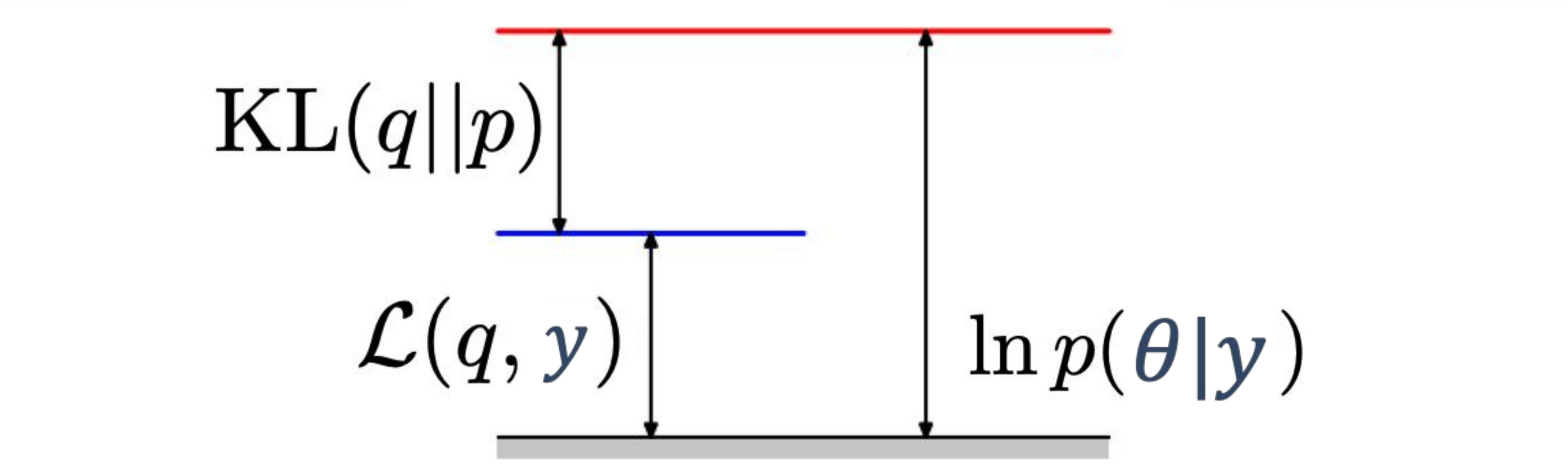
Variational Inference

$$p(\theta|y) = \frac{p(y|\theta)p(\theta}{\int p(y,\theta)d}$$

Kullback-Leibler divergence:

$$p(\theta|y) \approx q* = argmin_{q\in Q} f(q(\theta), p(\theta$$

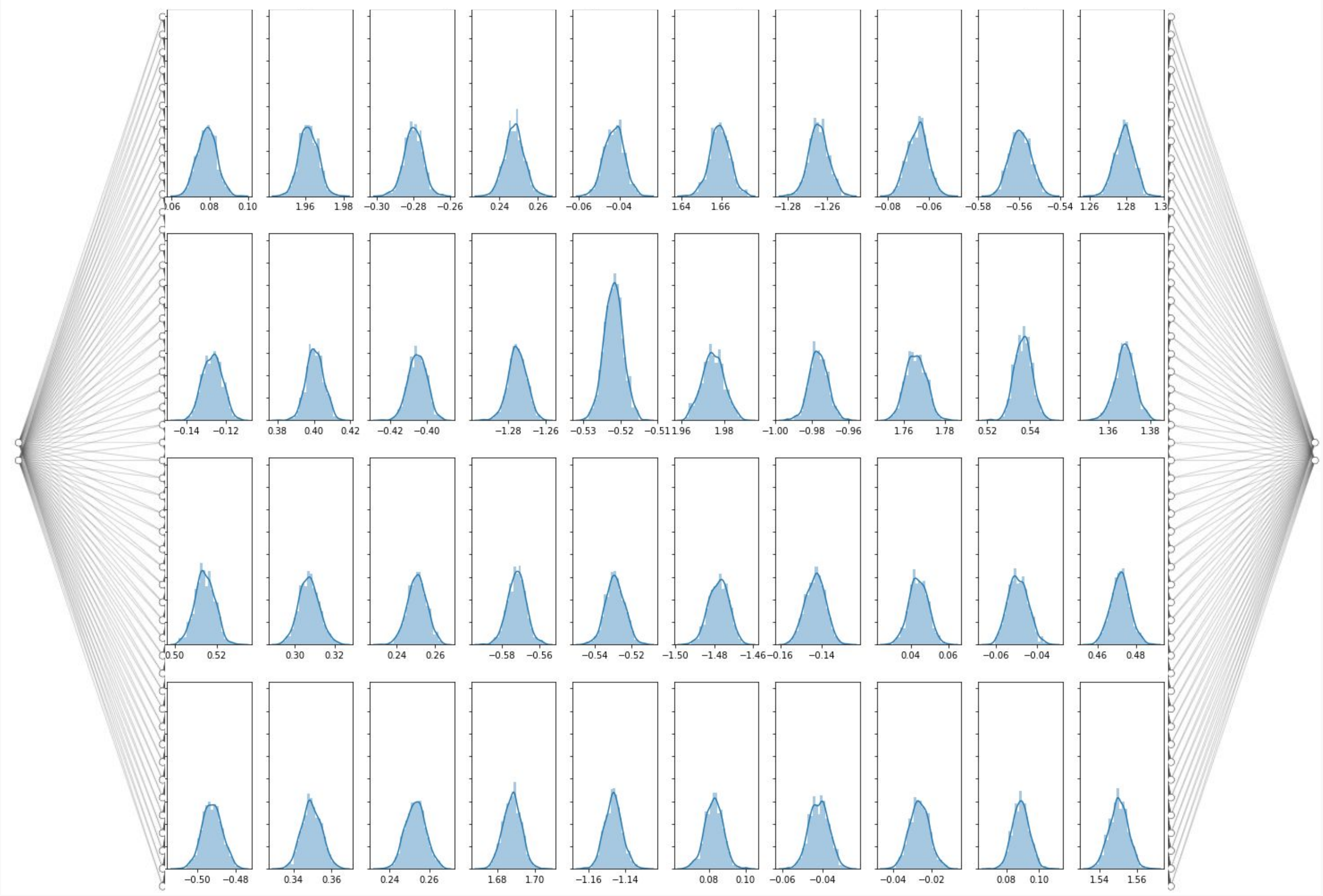
$$argmin_q KL(q,p) \equiv argmax_q EL$$



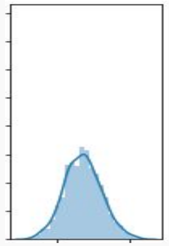
Bishop

Variational Inference

Input



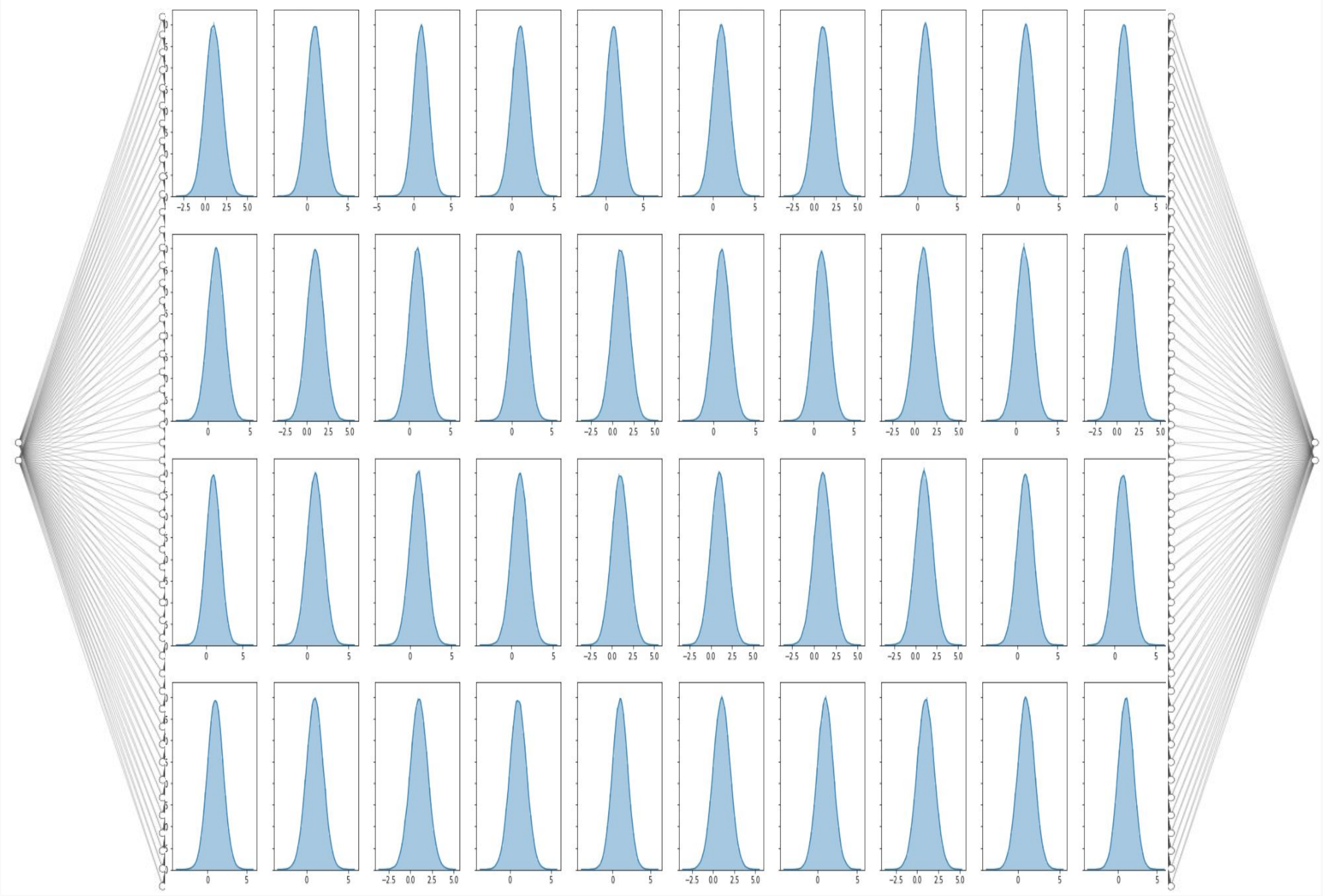
Output



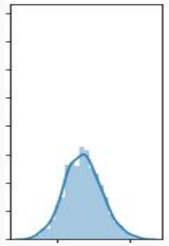
Variational Inference

$$p(\theta|y) \approx q *= argmax_{q \in Q} ELB$$

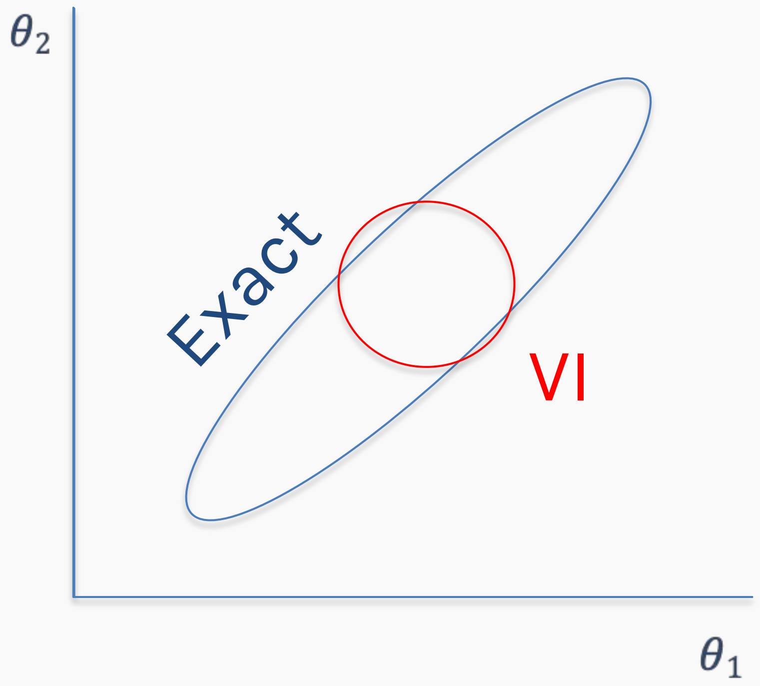
Input



Output



Variational Inference



$$KL(q\,||\,p(\cdot\,|\,x)) = \int_\theta q(\theta)\log\frac{q(\theta)}{p(\theta\,|\,x)}d\theta$$

$$q(\theta) = \prod_{j=1}^{J} q_j(\theta_j)$$

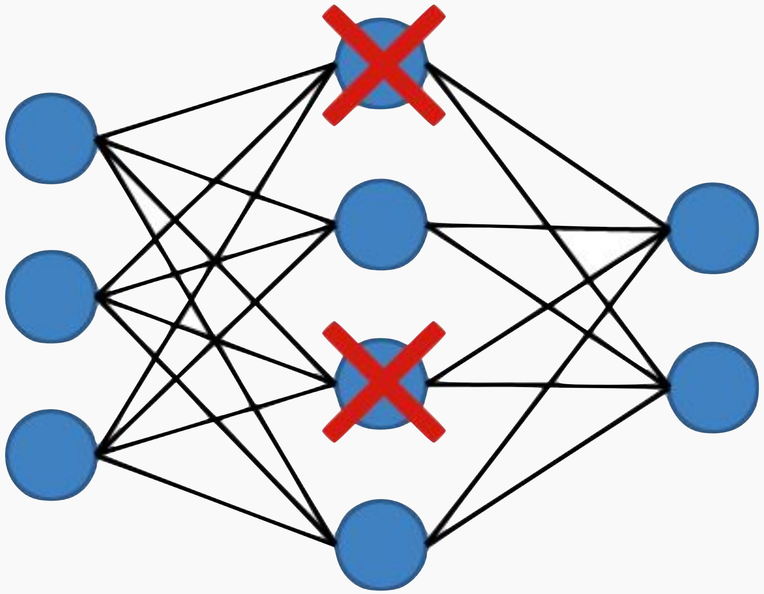
Underestimates variance (sometimes severely)

Dropout

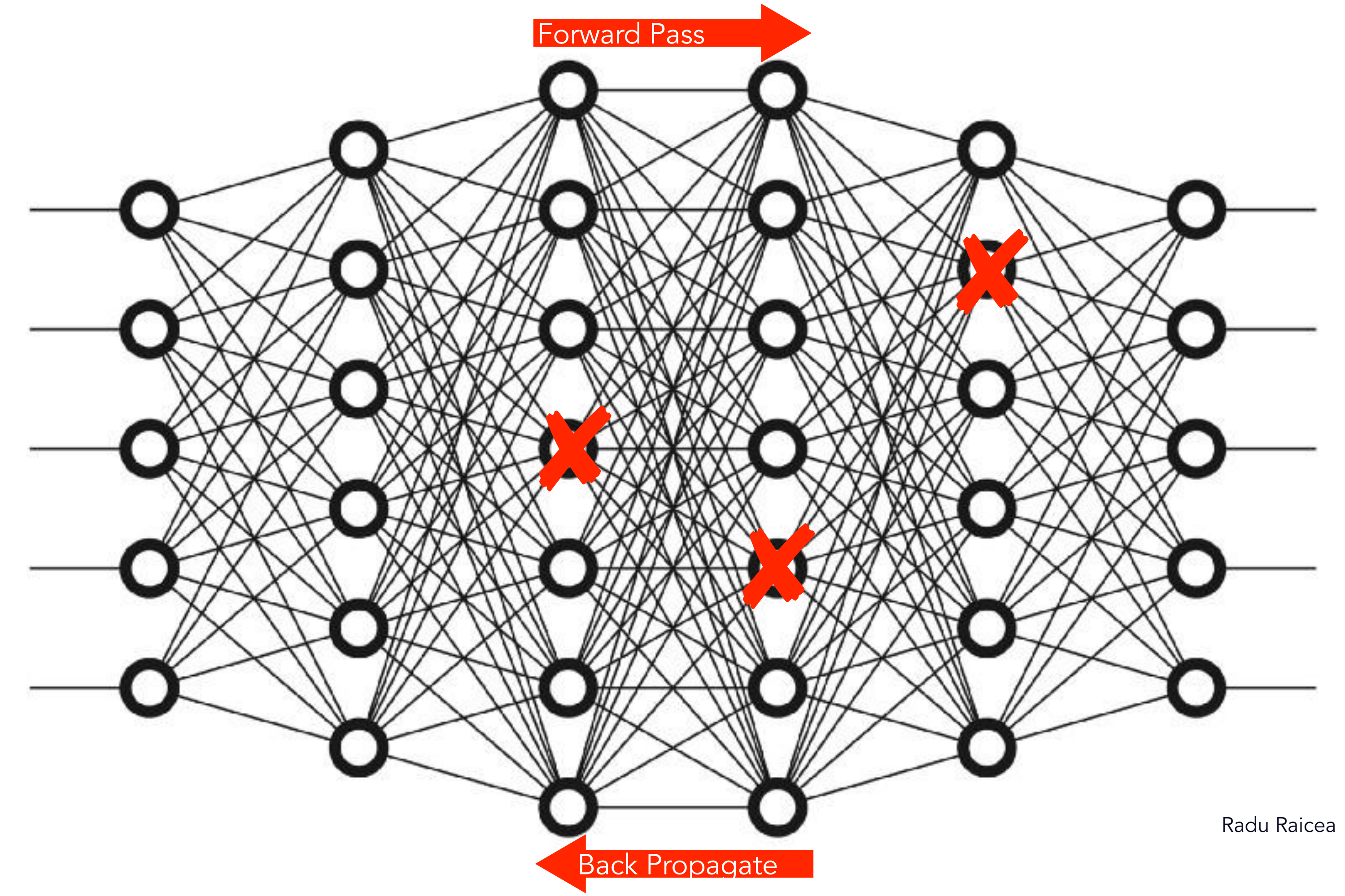
Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning

Yarin Gal YG279@CAM.AC.UK
Zoubin Ghahramani ZG201@CAM.AC.UK
University of Cambridge

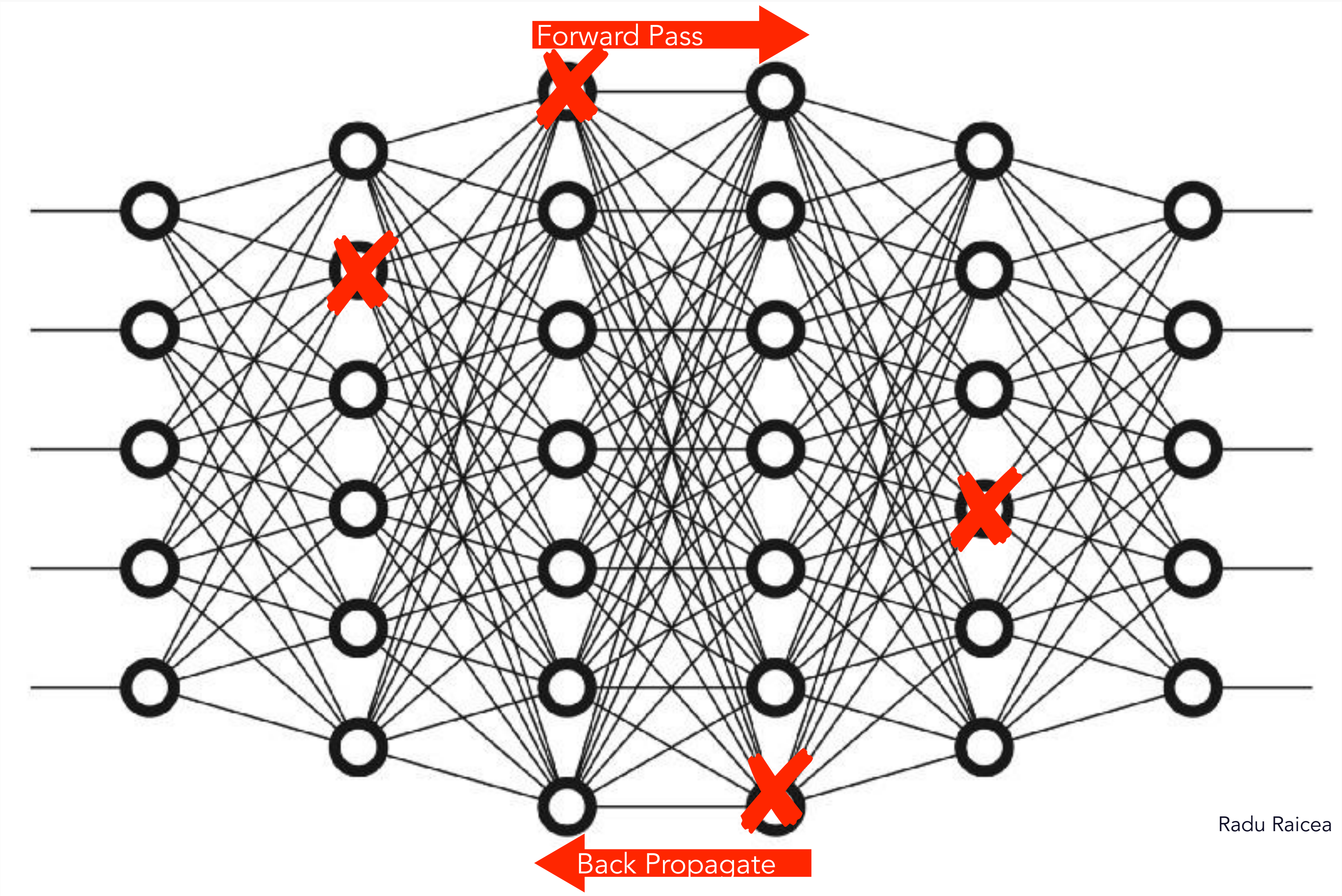
arXiv:1506.02142



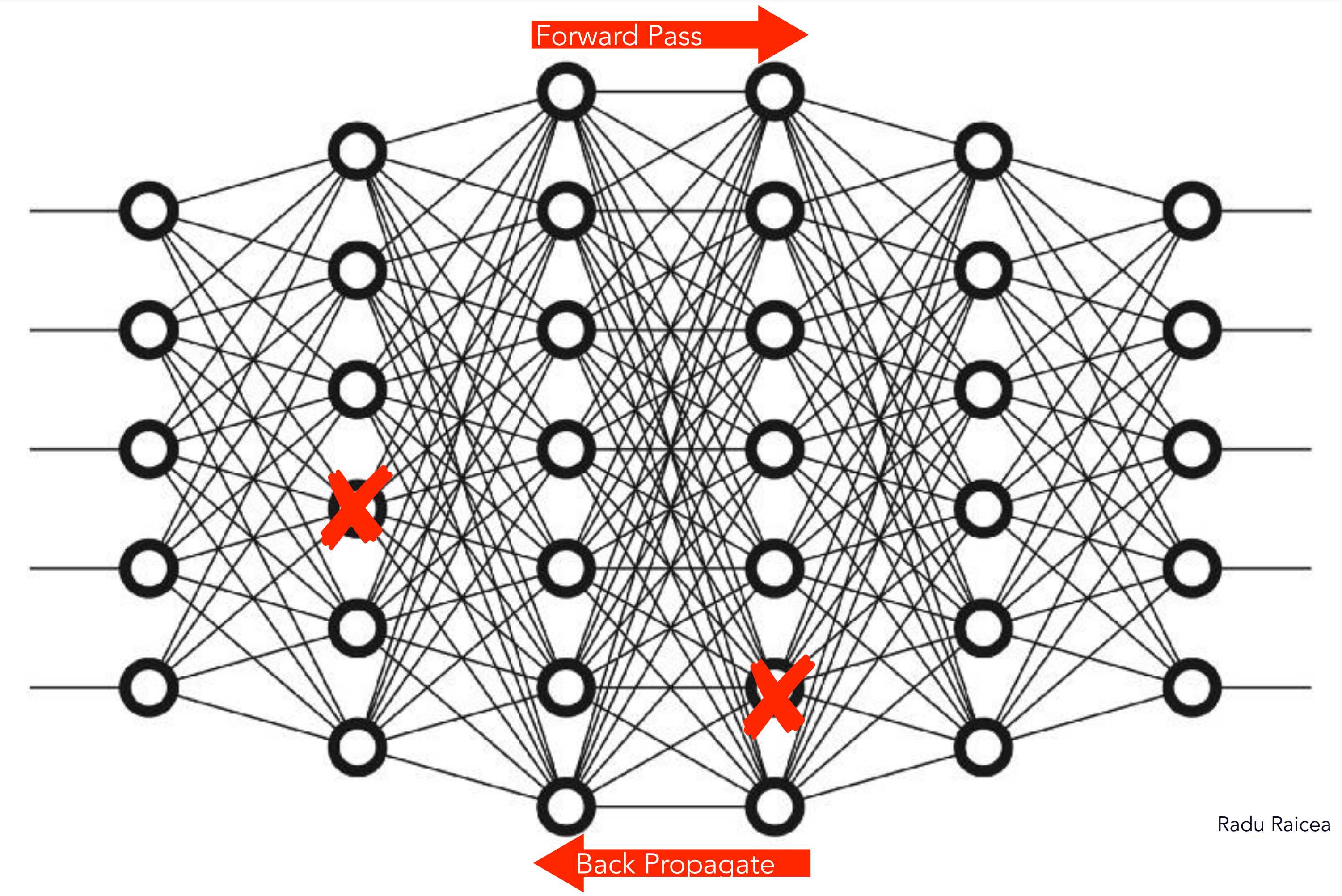
Dropout: train



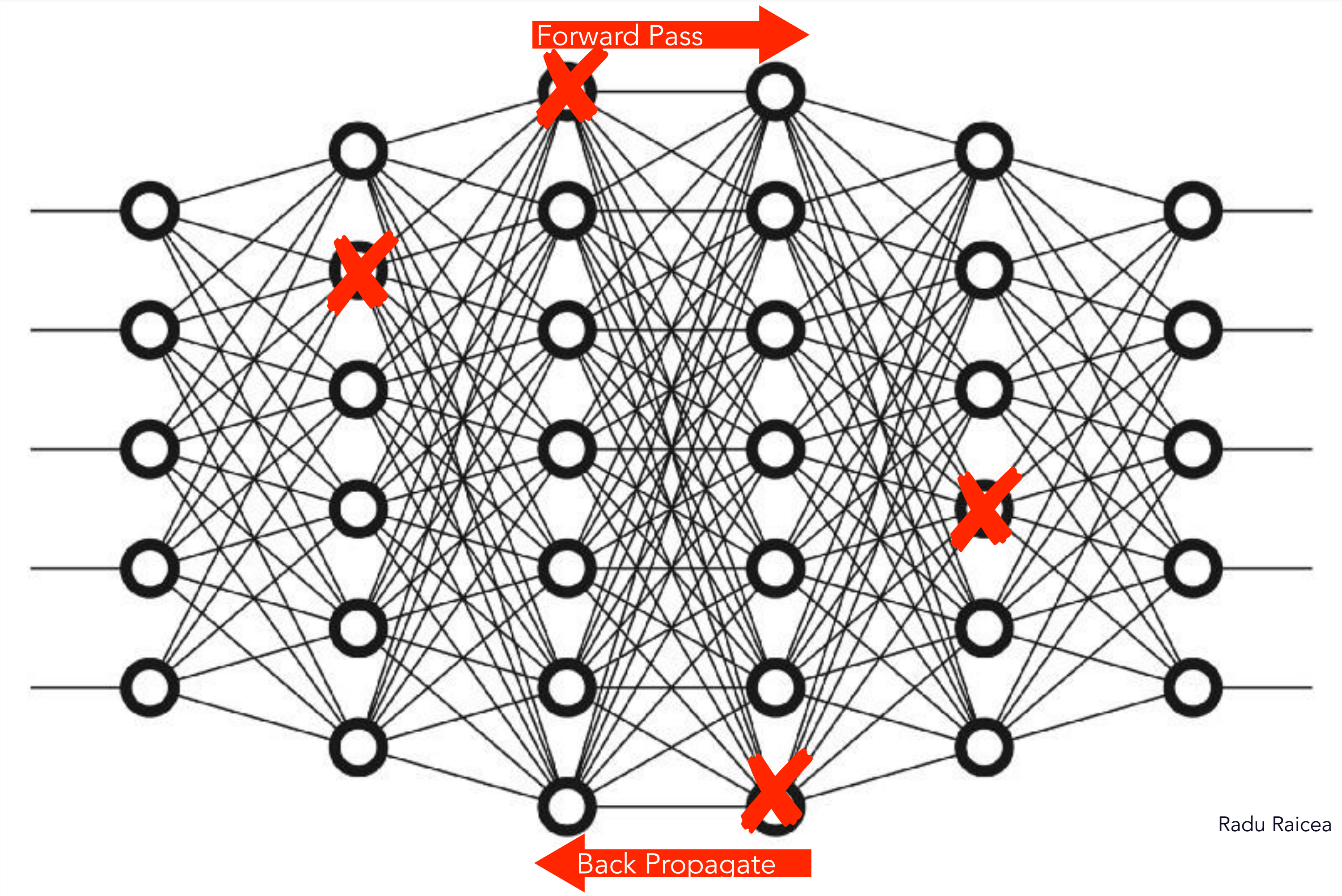
Dropout: train



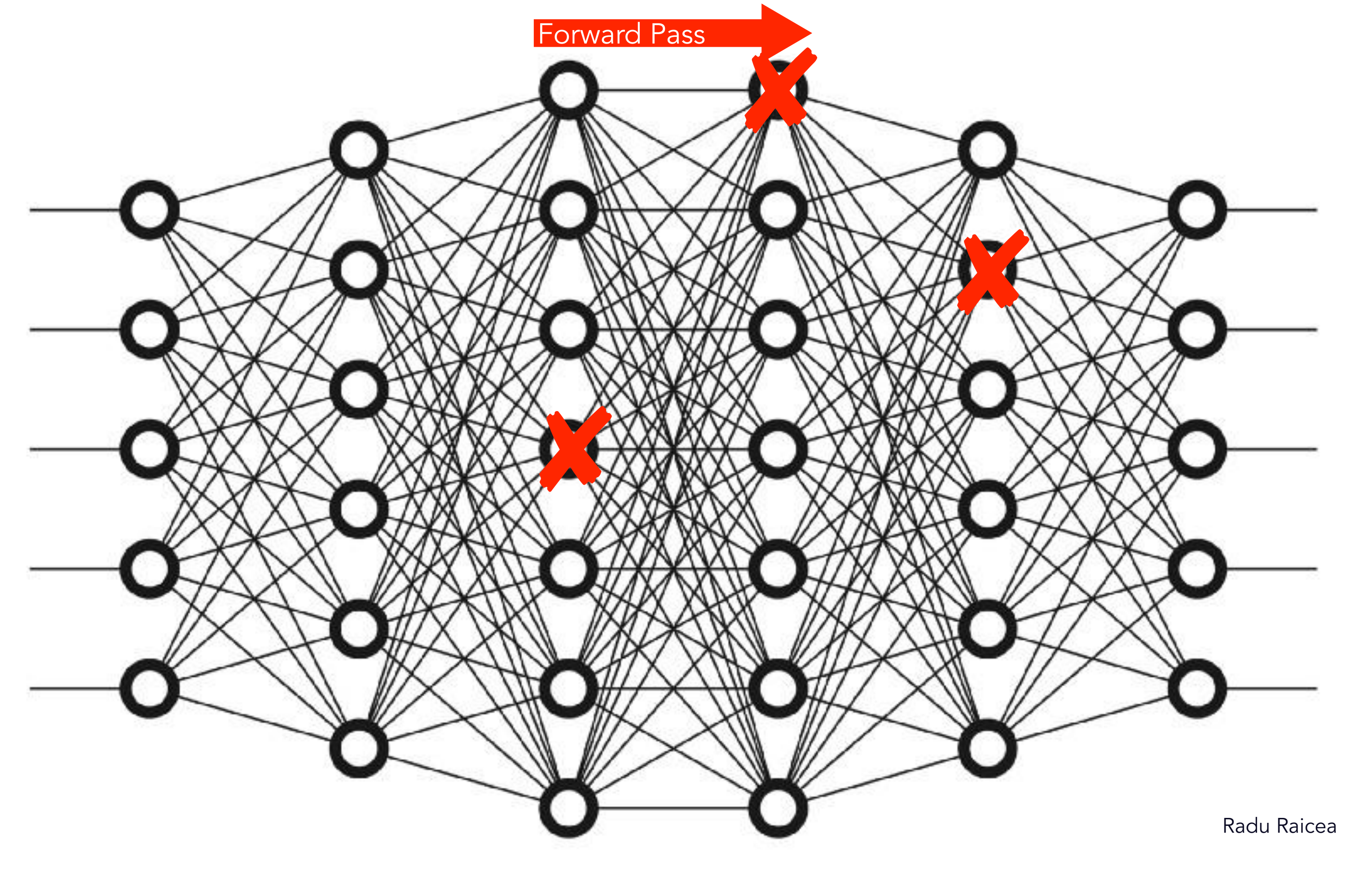
Dropout: train



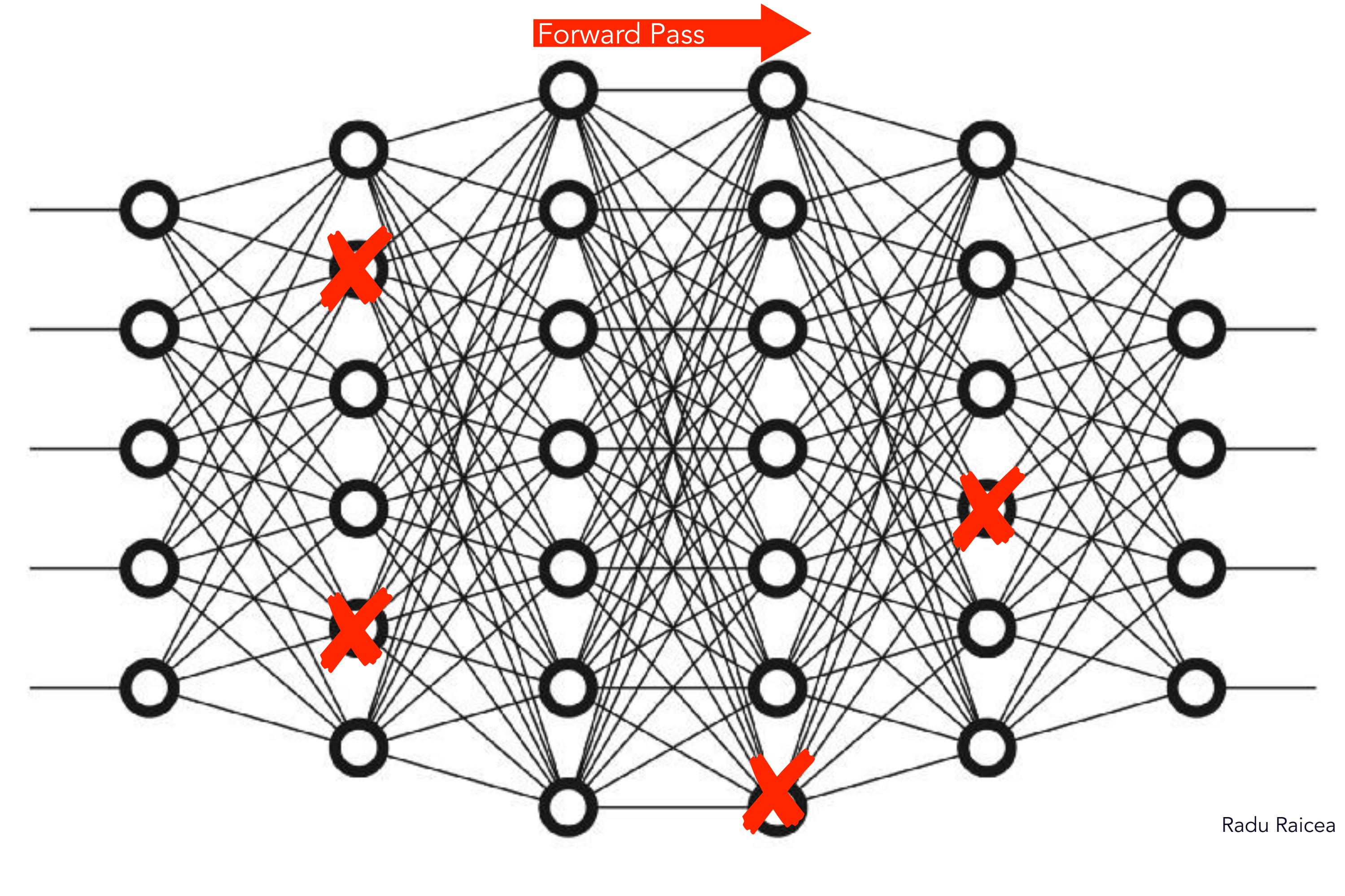
Dropout: train

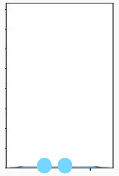


Dropout: evaluate

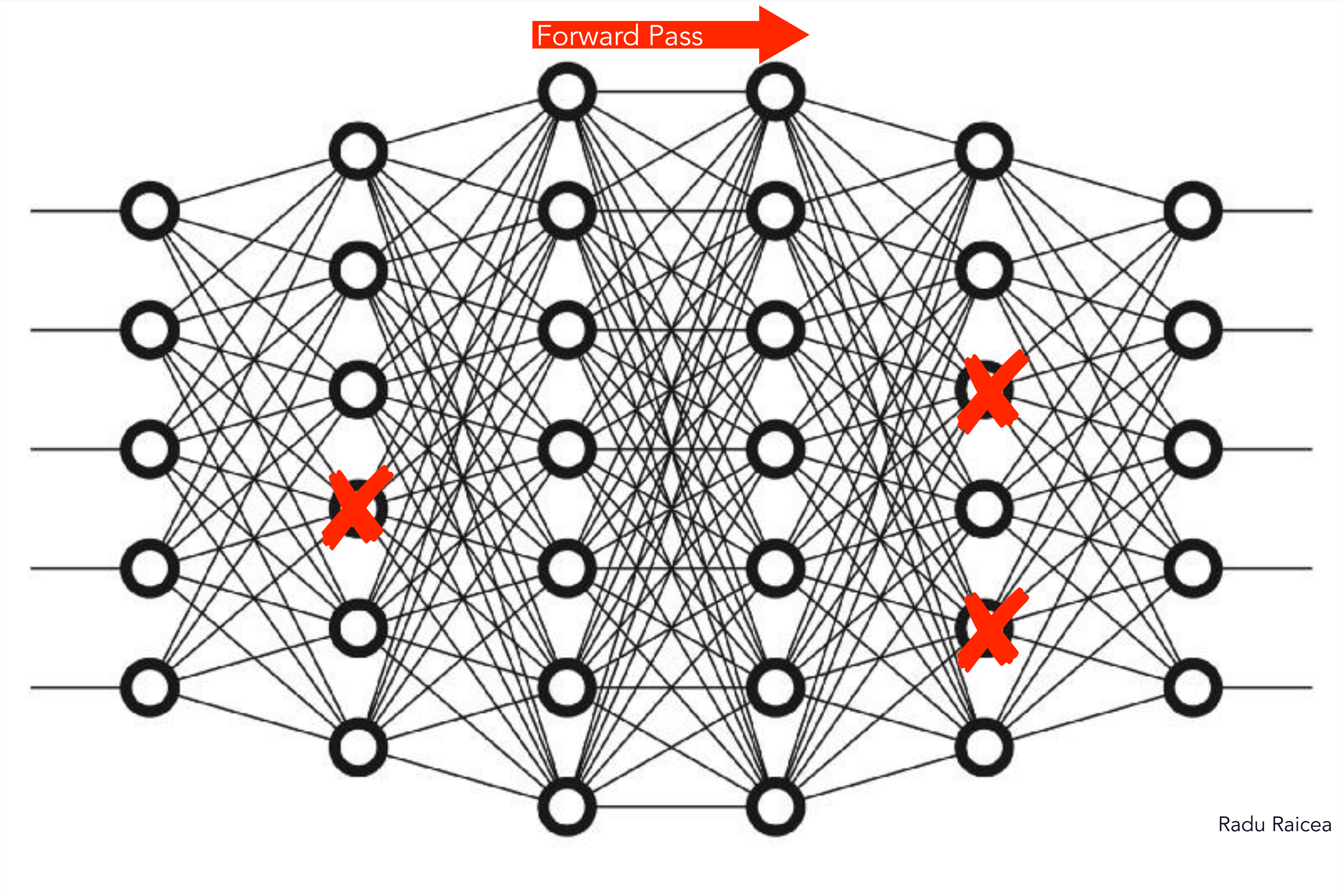


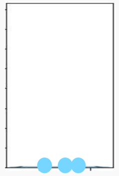
Dropout: evaluate



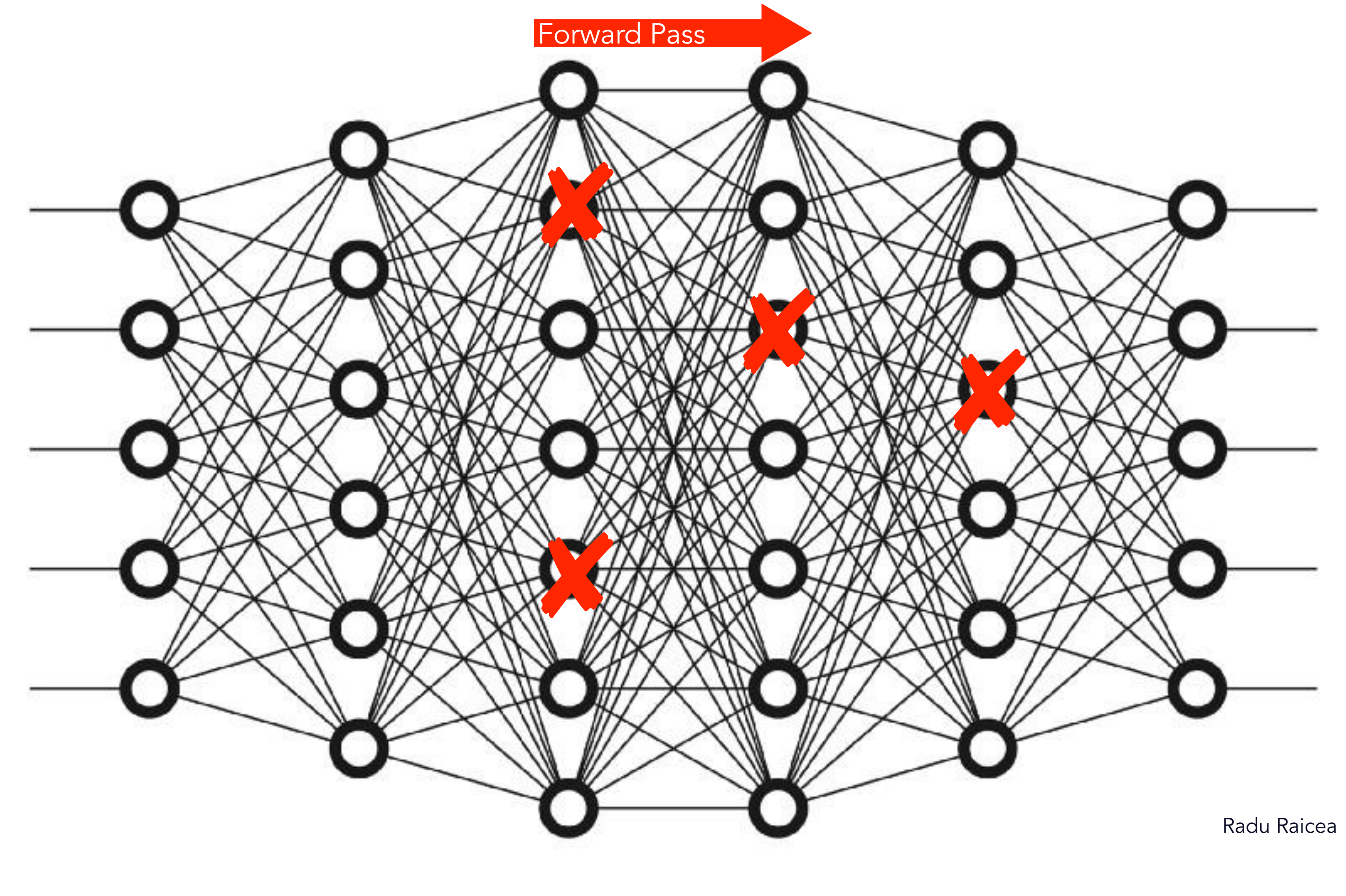


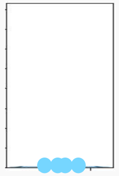
Dropout: evaluate



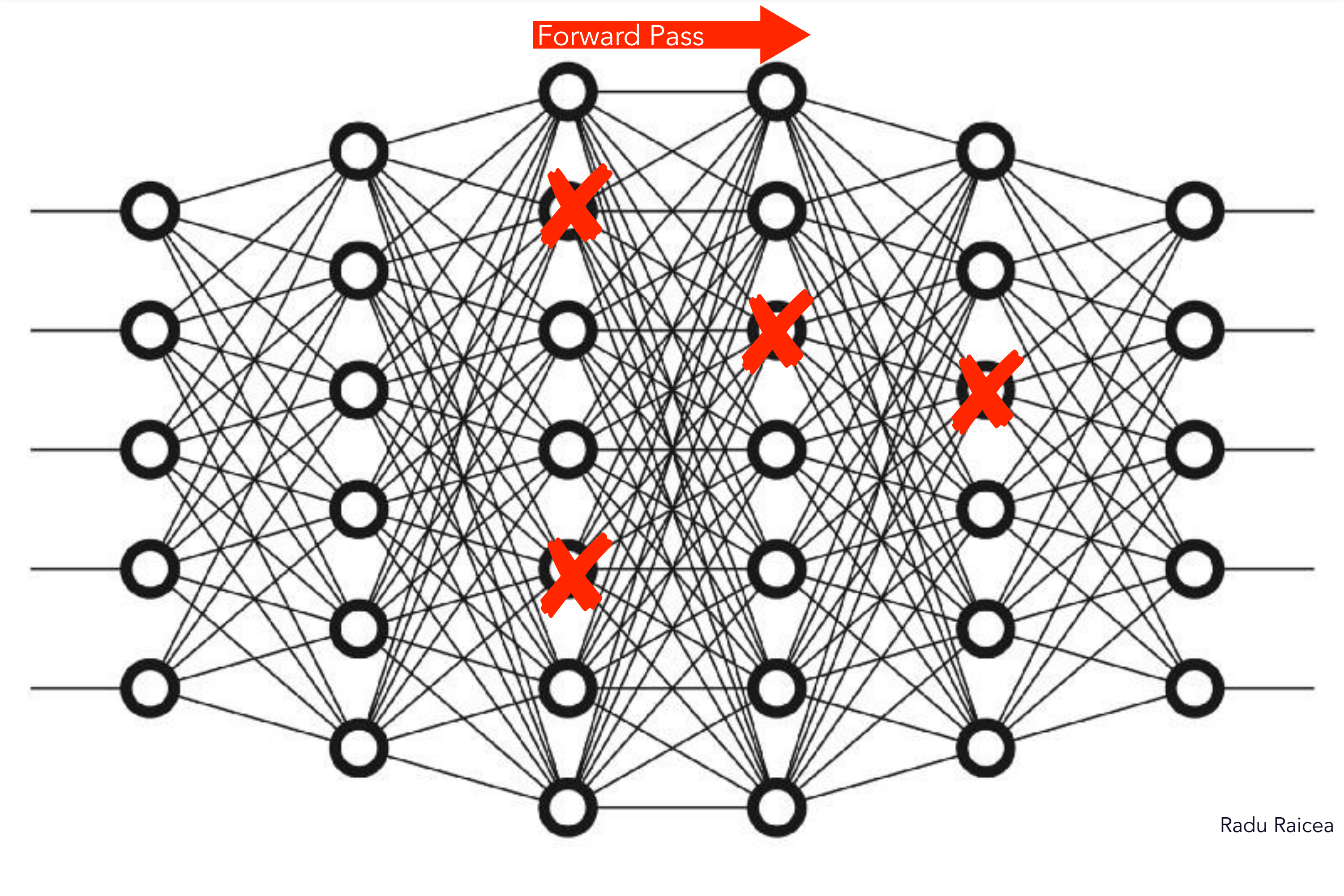


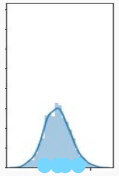
Dropout: evaluate



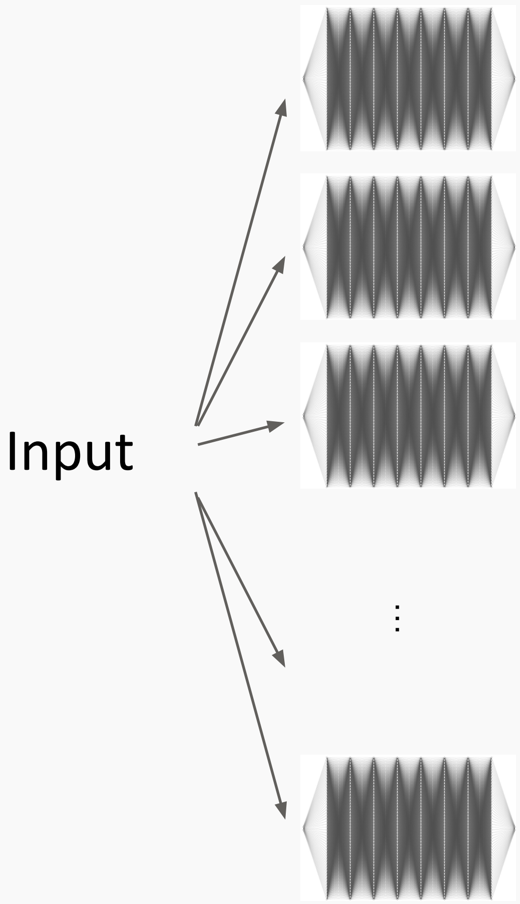


Dropout: evaluate

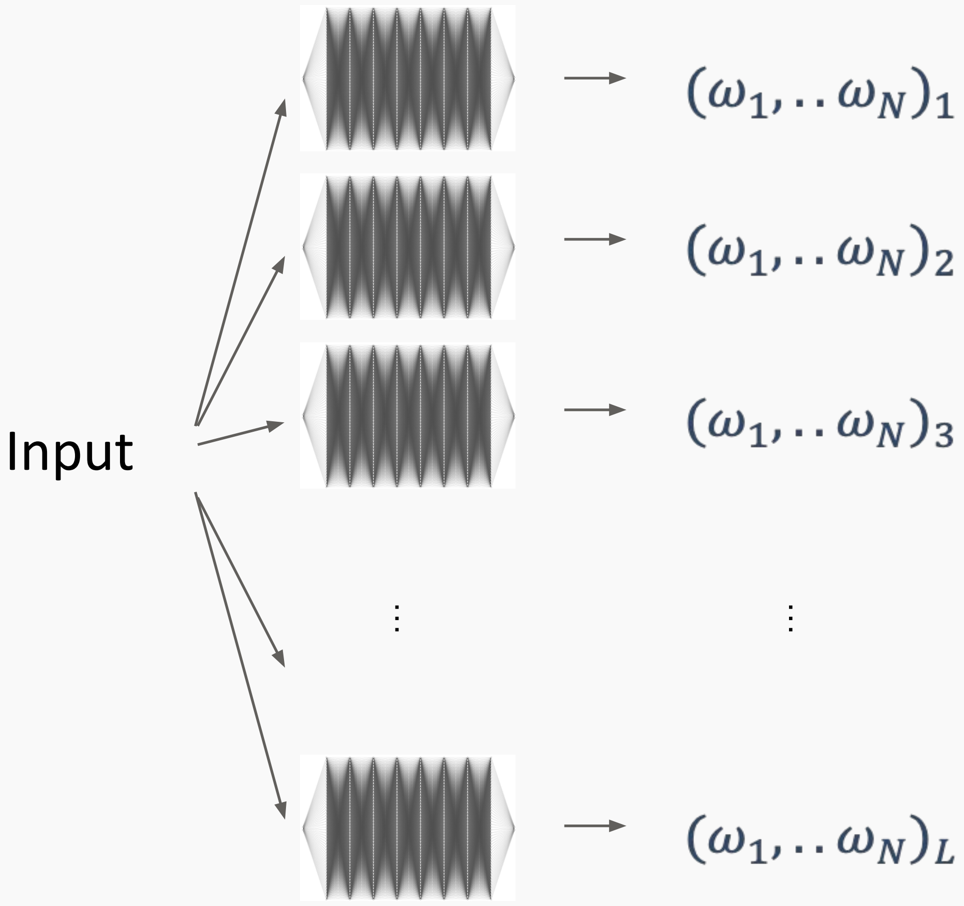




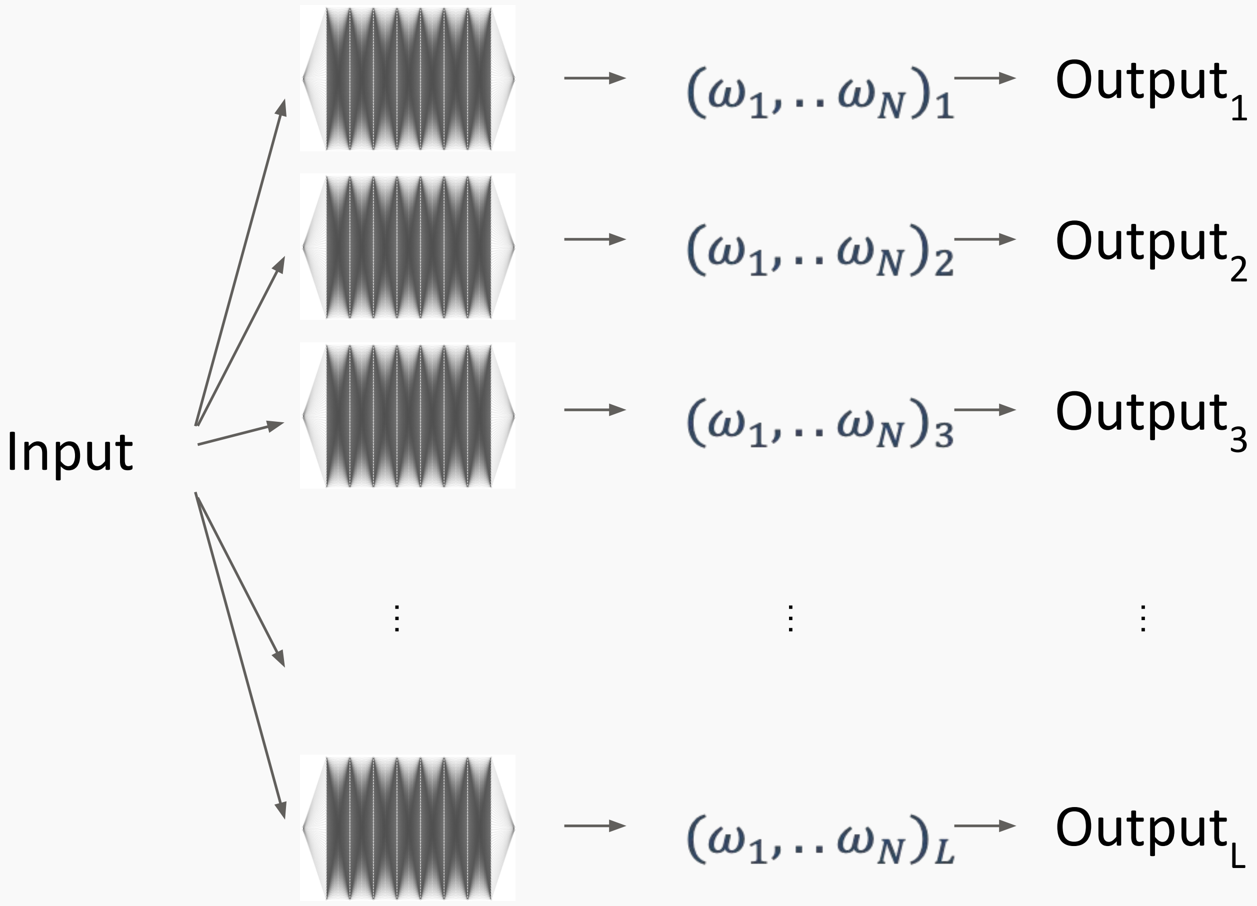
Bootstrap for Inference



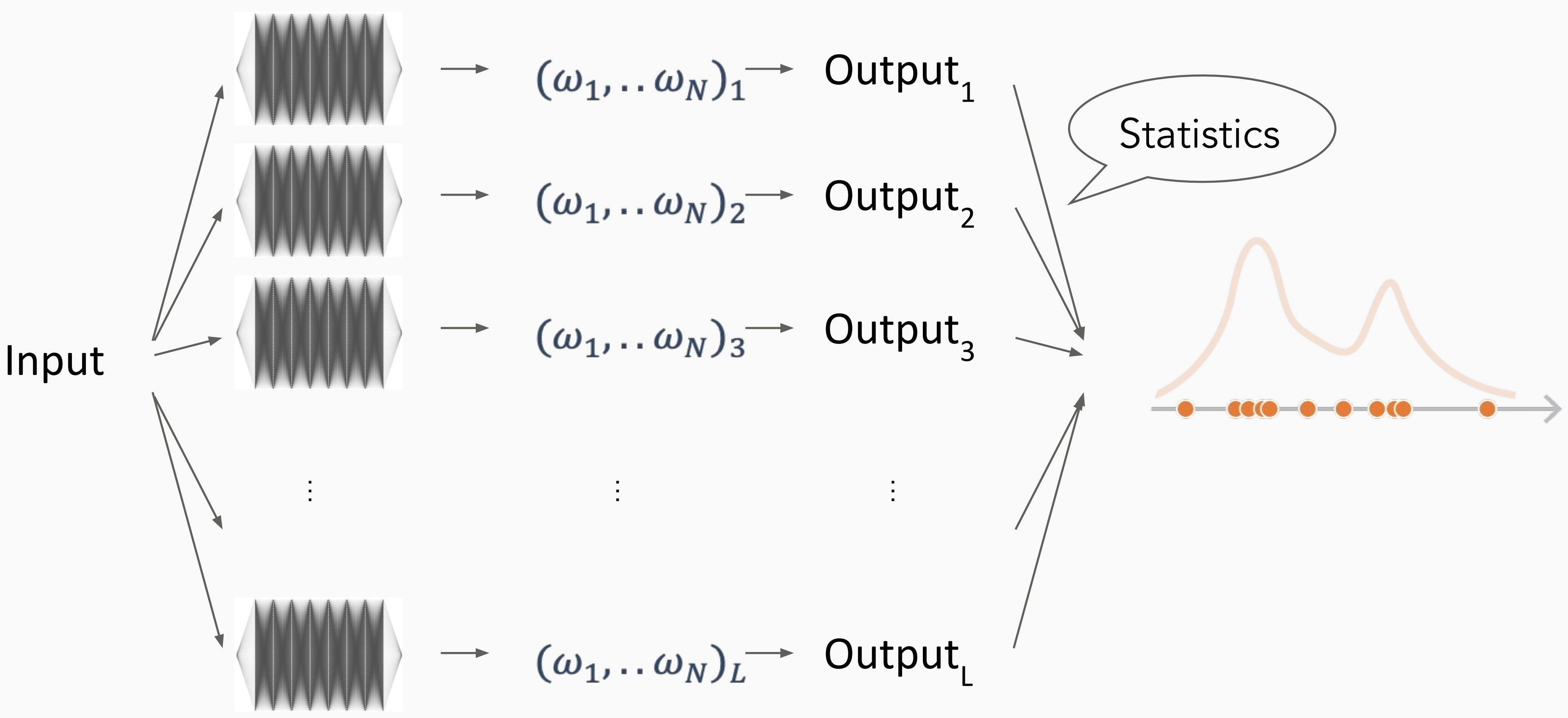
Bootstrap for Inference



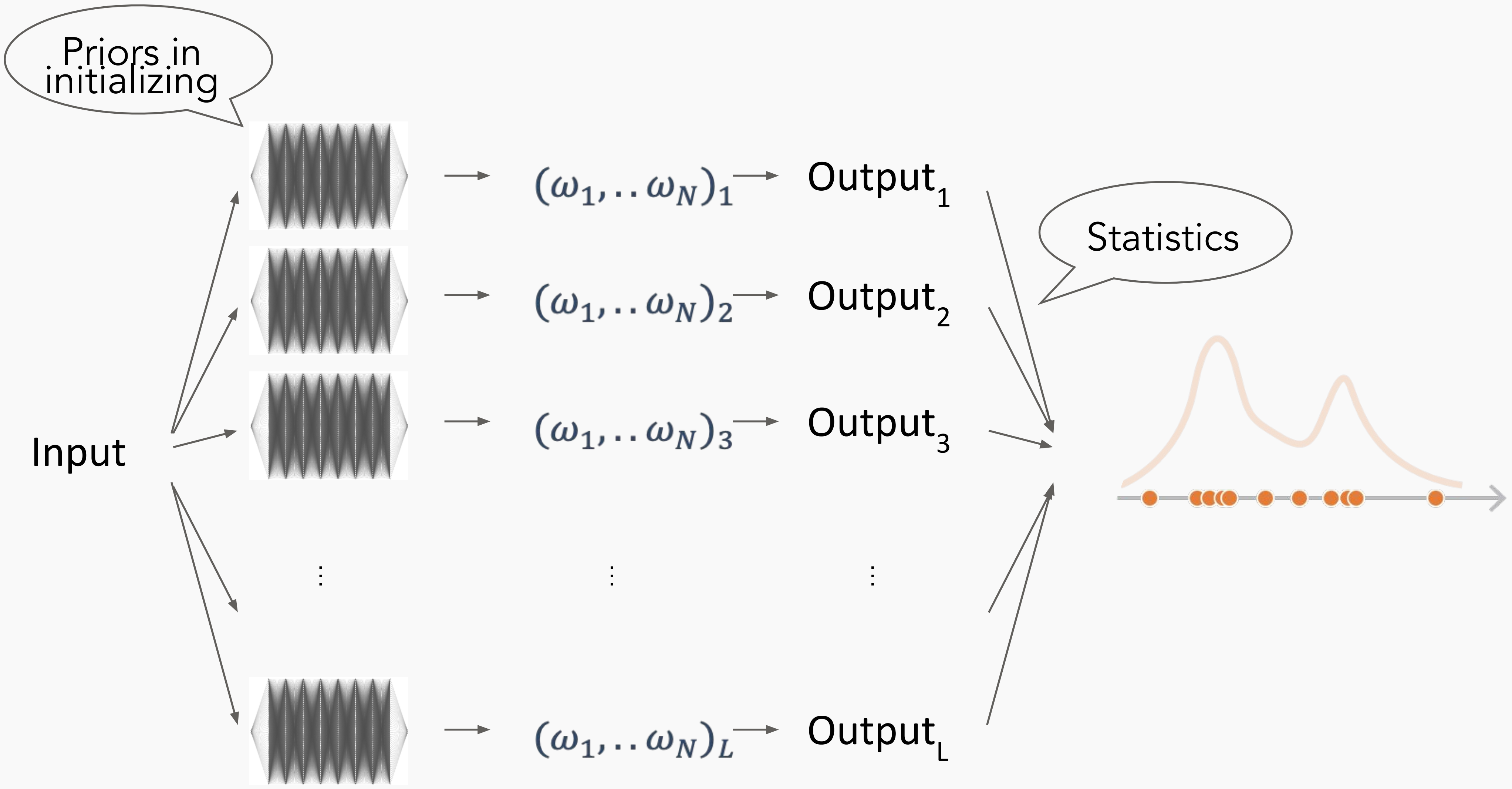
Bootstrap for Inference



Bootstrap for Inference



Bootstrap for Inference



Working Example

Variational Bayesian Inference

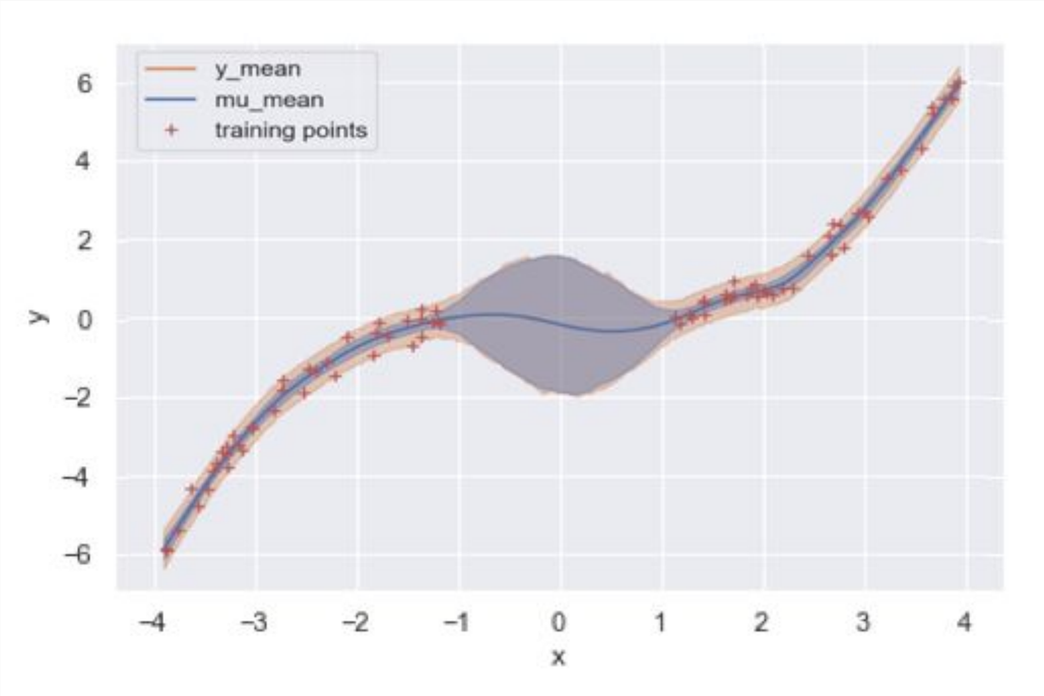
The problem

$$y = x^3 + \mathrm{N}(0,0.25)$$

Working Example

Variational Bayesian Inference

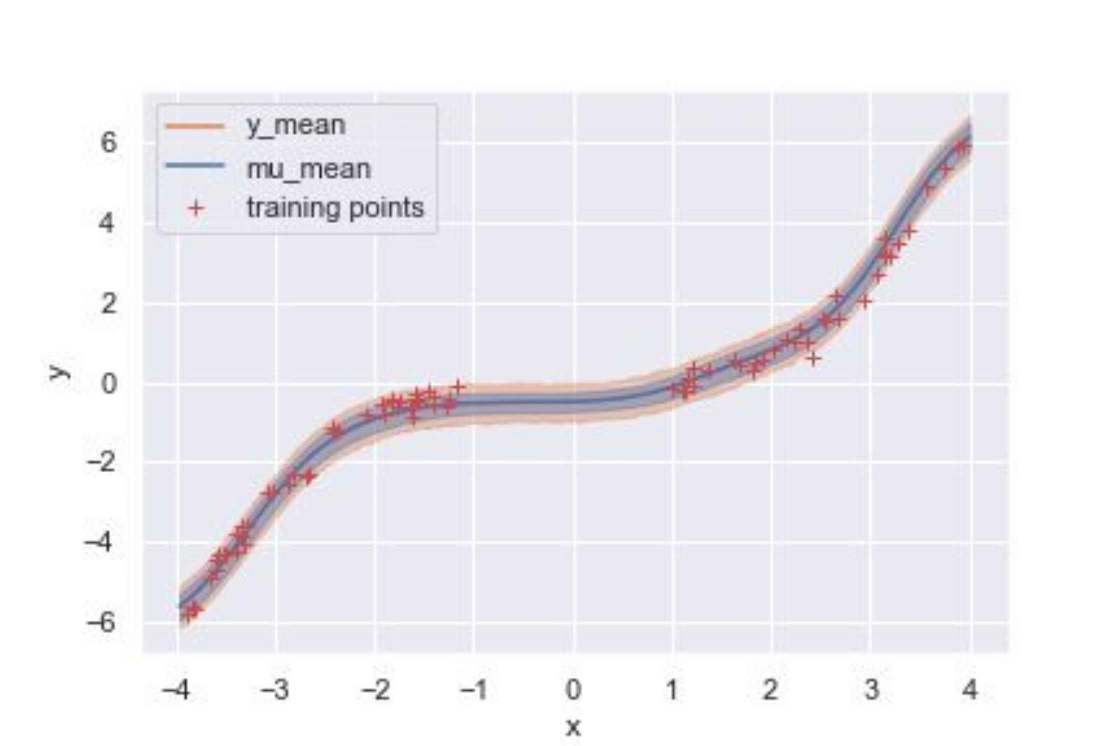
The right solution (MCMC)



Working Example

Variational Bayesian Inference

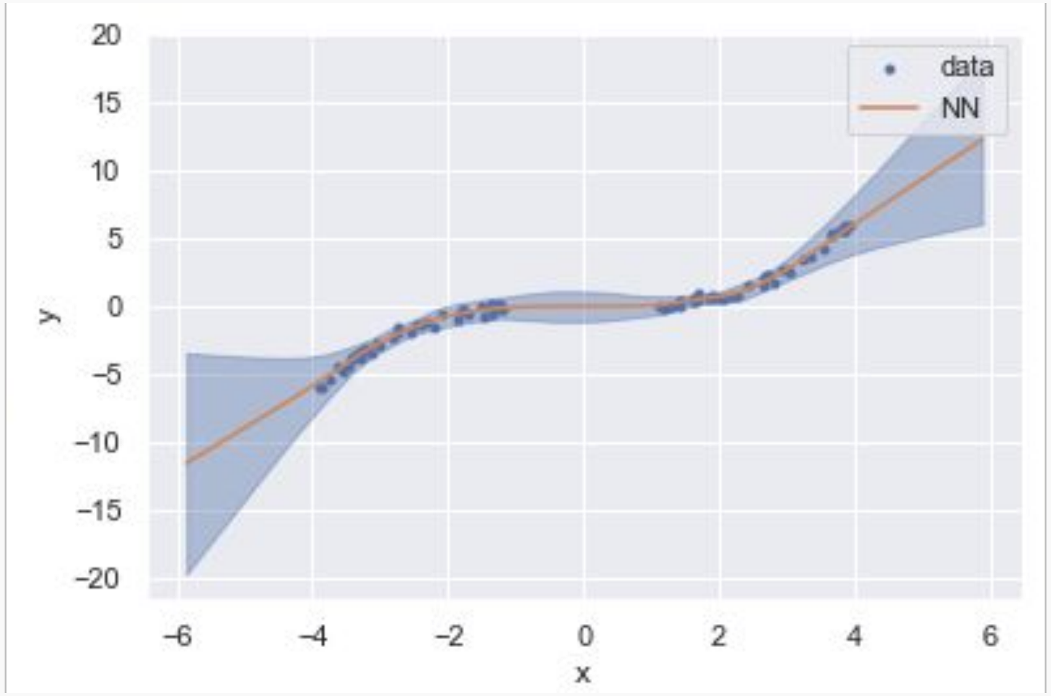
SVI



Working Example

Variational Bayesian Inference
Bootstrap

Model Mean 95% models

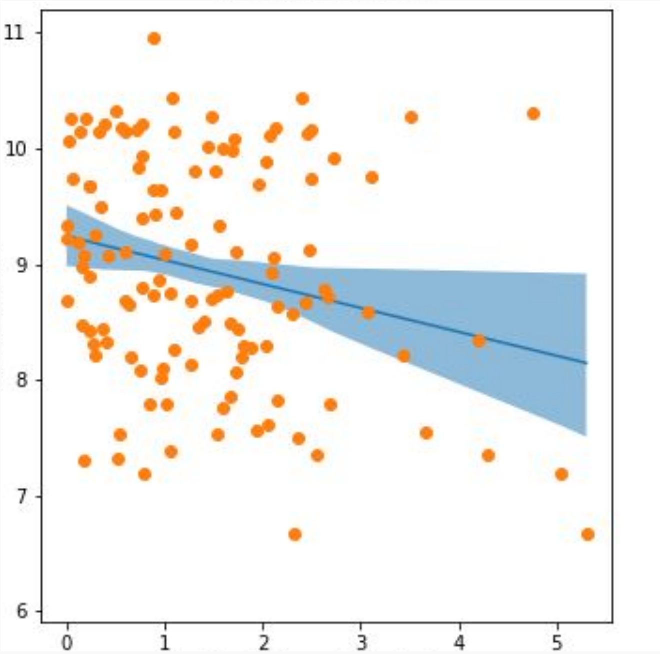


Working Example

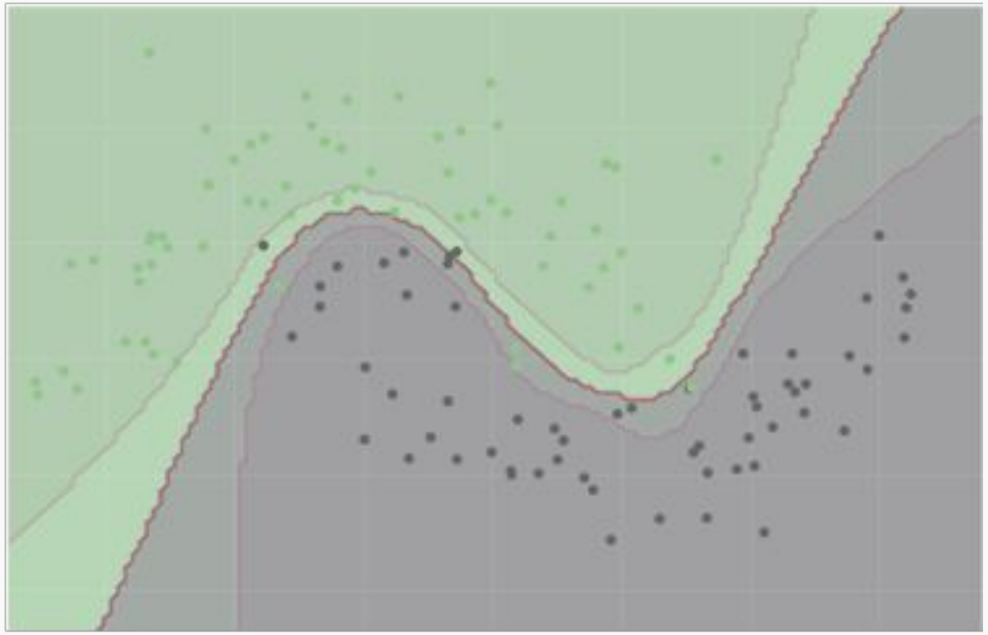
Variational Bayesian Inference
Bootstrap

Model Mean 95% models

Regression



Classification

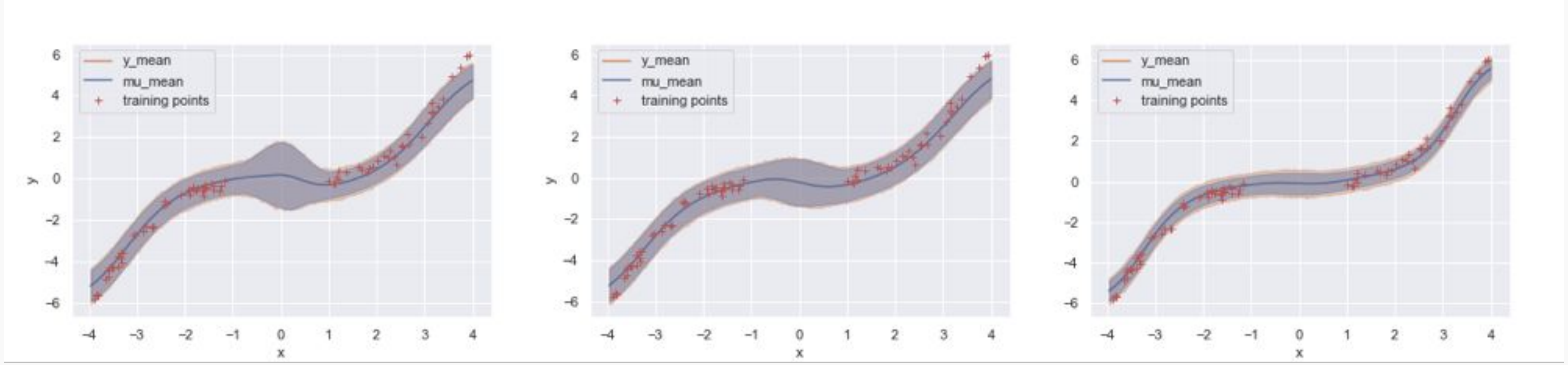


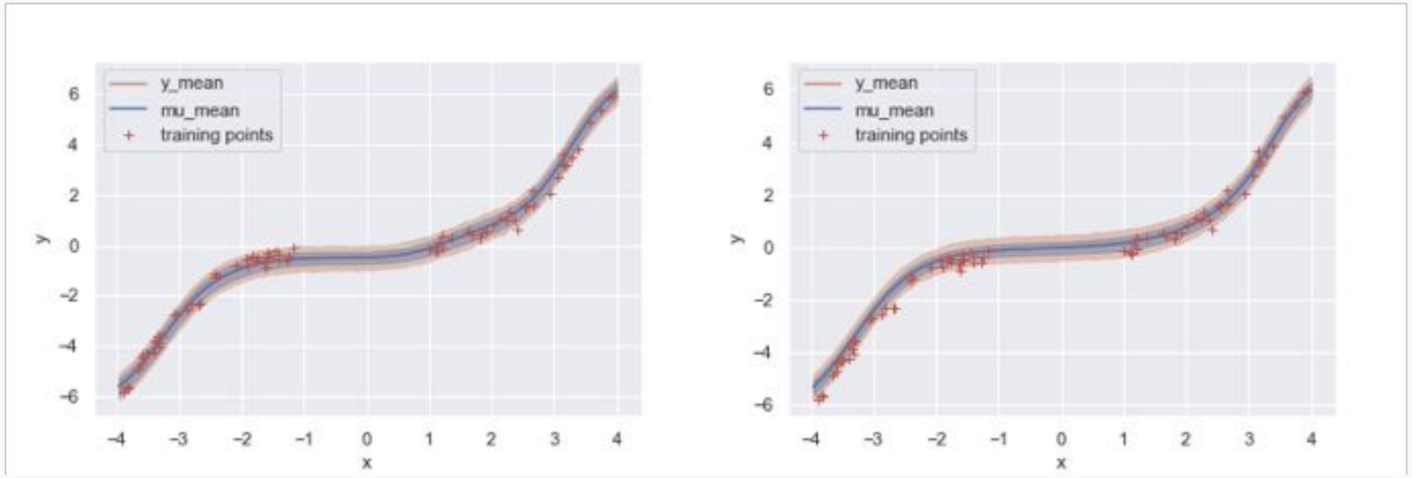
Extra

Working Example

Variational Bayesian Inference

SVI





DONE