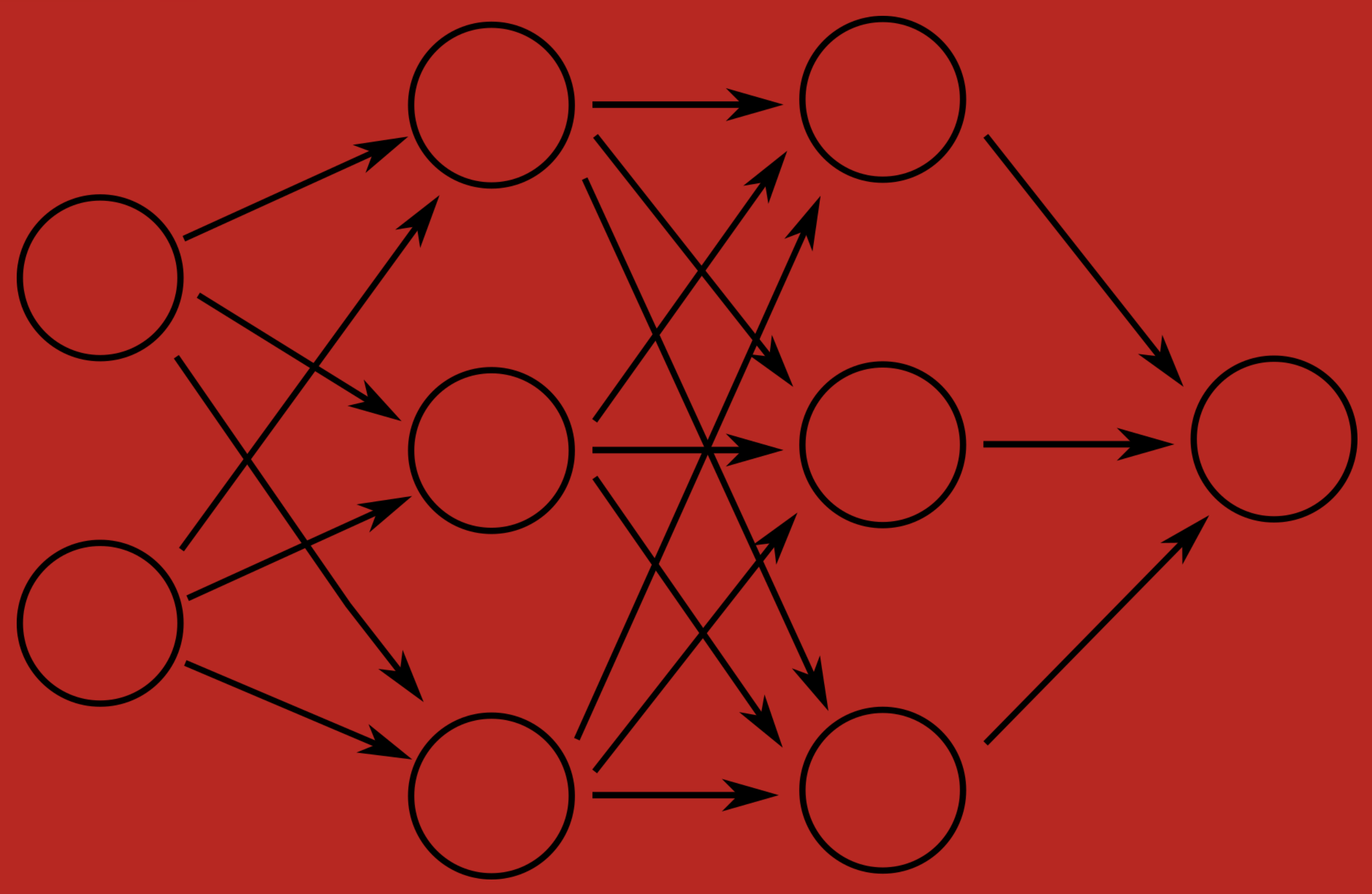


Uncertainty Quantification in Neural Networks: Methods and Considerations

Laurens Sluijterman - Mathematics department
Eric Cator, Tom Heskes

Outline

Outline

Main point: How to evaluate a method

Outline

Main point: How to evaluate a method

Outline

Main point: How to evaluate a method

1. Why are neural networks uncertain?

Outline

Main point: How to evaluate a method

1. Why are neural networks uncertain?
2. Methods: MVE, ensembling, dropout

Outline

Main point: How to evaluate a method

1. Why are neural networks uncertain?
2. Methods: MVE, ensembling, dropout
3. Out of distribution

Terminology

Terminology

Terminology

Systematic / Statistical

Terminology

Systematic / Statistical

Terminology

Systematic / Statistical

Epistemic / Aleatoric

Terminology

Systematic / Statistical

Epistemic / Aleatoric

Terminology

Systematic / Statistical

Epistemic / Aleatoric

Why are we uncertain?

Why are we uncertain?

Why are we uncertain?

1. Uncertain about model parameters

Why are we uncertain?

1. Uncertain about model parameters
    - Training data is random

Why are we uncertain?

1. Uncertain about model parameters
    - Training data is random
    - Optimization is random

Why are we uncertain?

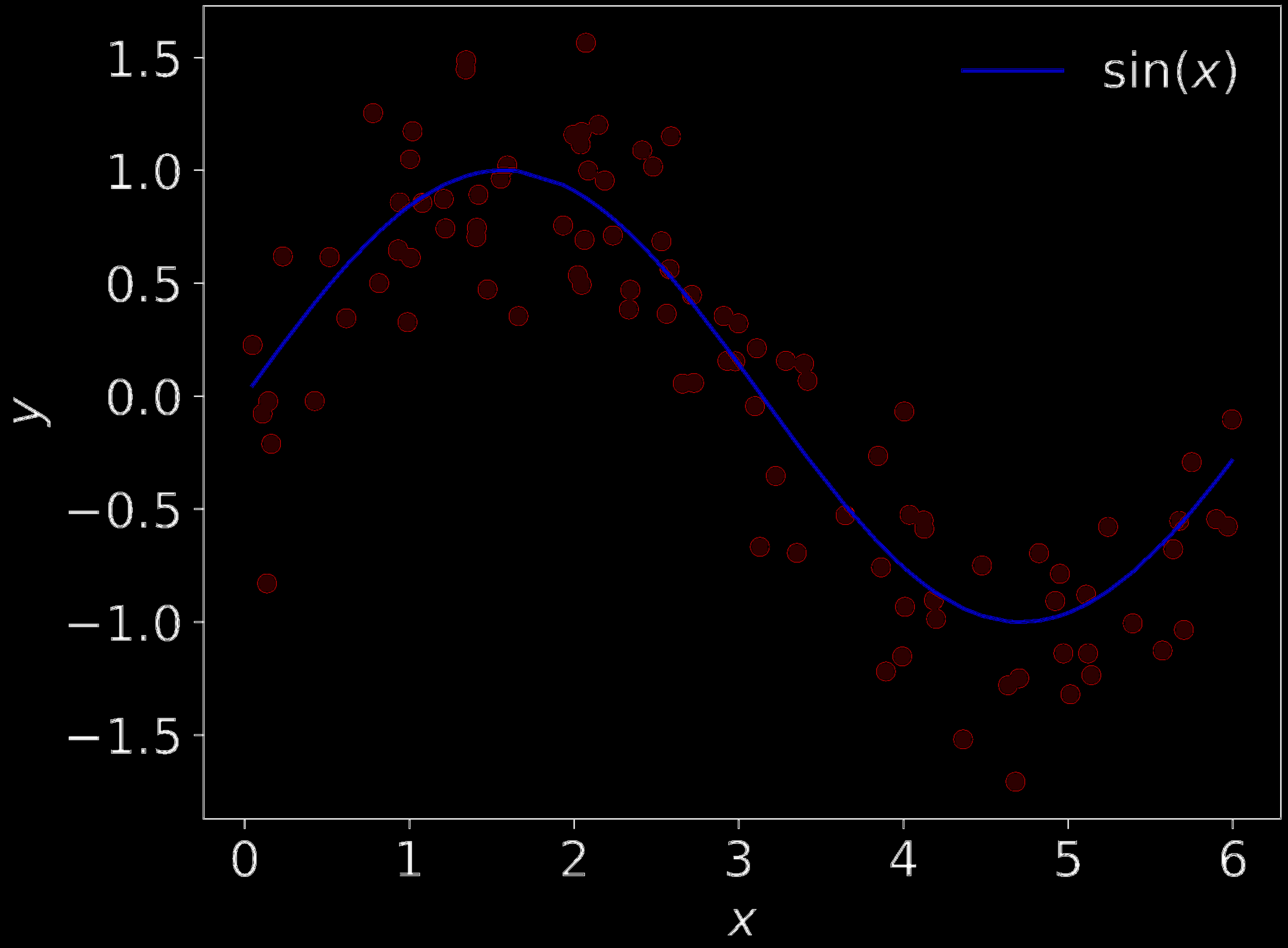
1. Uncertain about model parameters
    - Training data is random
    - Optimization is random
2. Outcome is random

Why are we uncertain?

1. Uncertain about model parameters
   - Training data is random
   - Optimization is random
2. Outcome is random
3. Uncertain about model

Why are we uncertain?

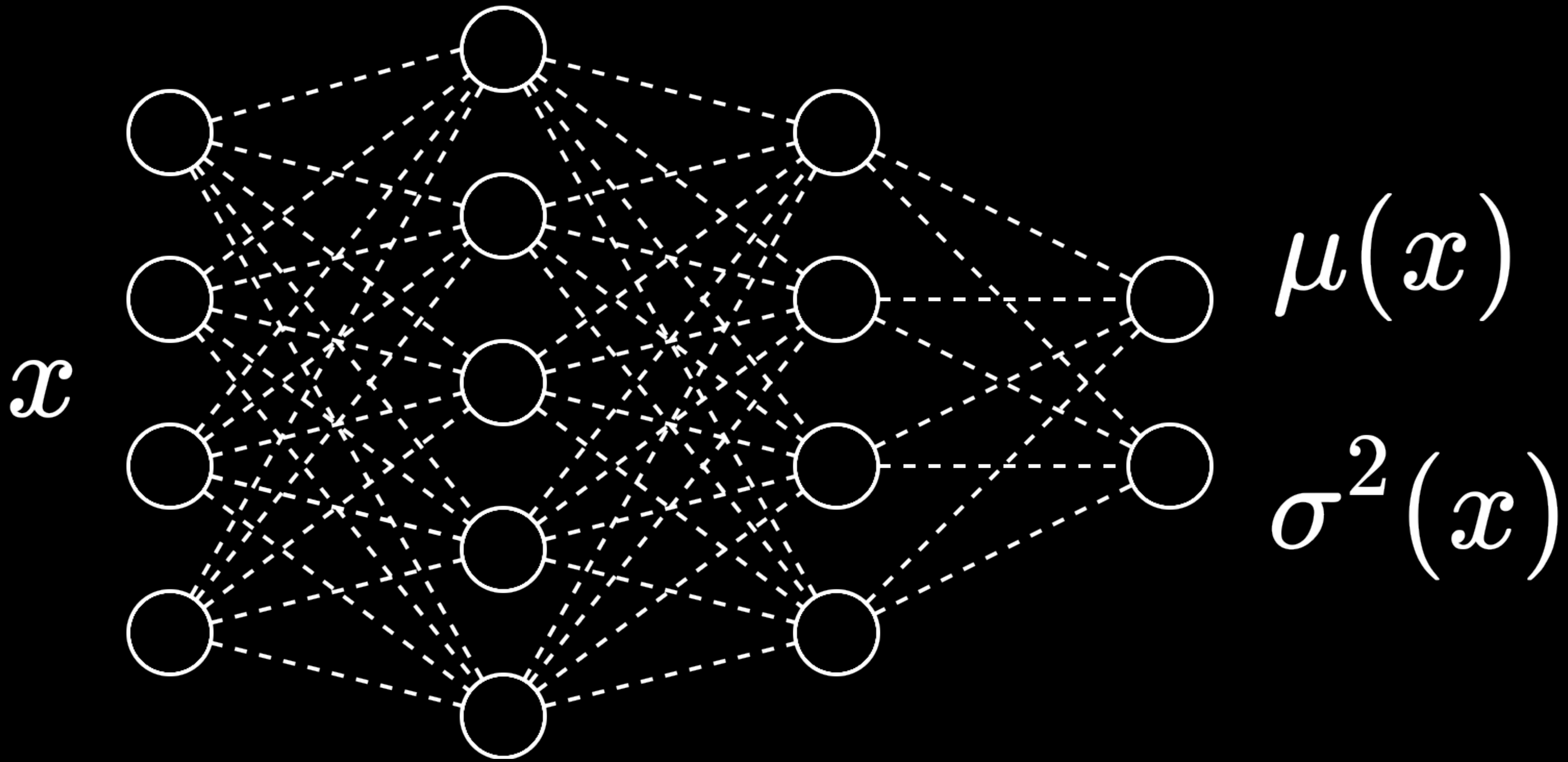
1. Uncertain about model parameters
   - Training data is random
   - Optimization is random
2. Outcome is random
3. Uncertain about model



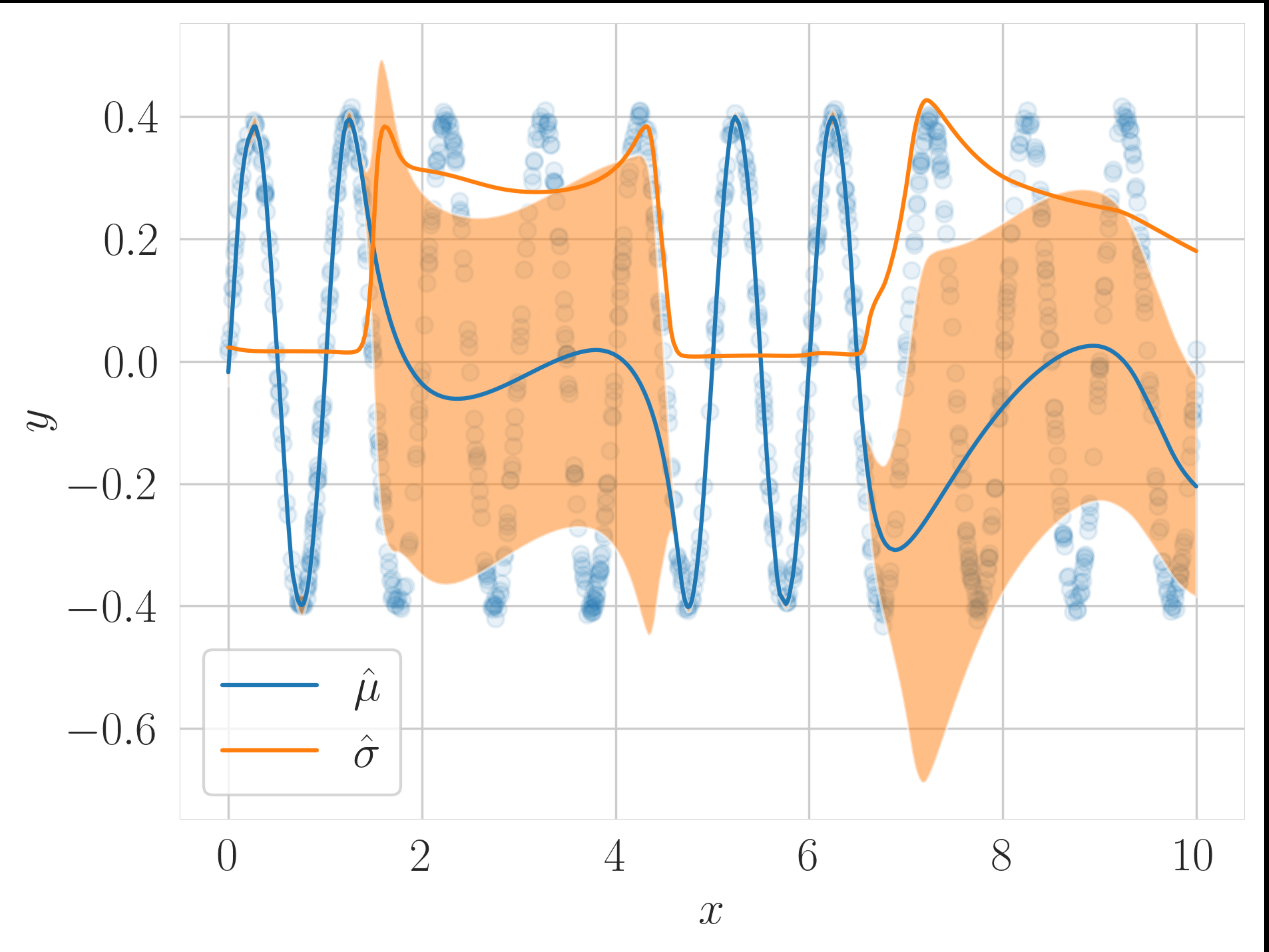
sin(x)
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
y
0
1
2
3
4
5
6
x

Learn the variance

Learn the variance

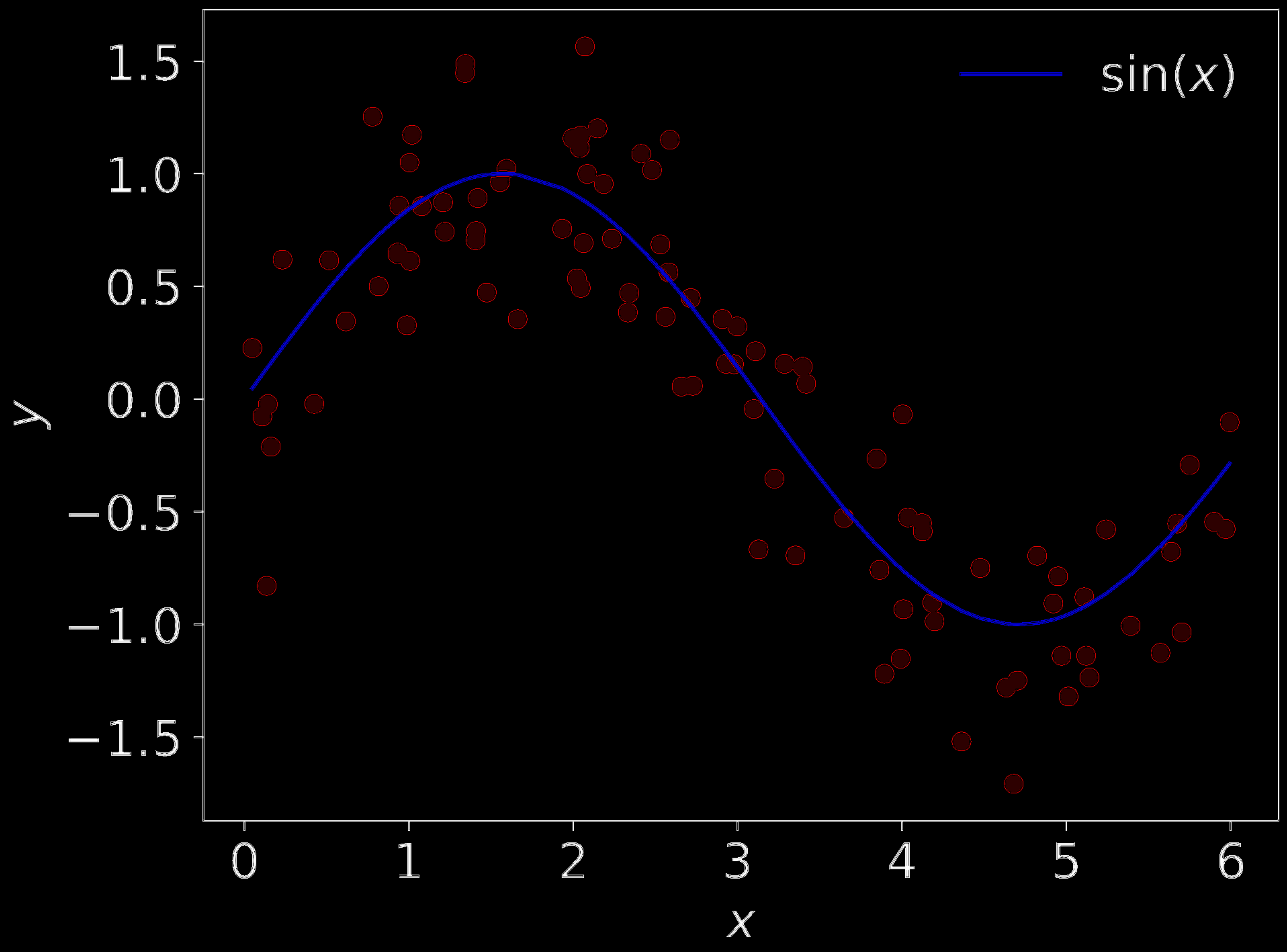


Sluijterman, Laurens, Eric Cator, and Tom Heskes. "Optimal Training of Mean Variance Estimation Neural Networks."



Sluijterman, Laurens, Eric Cator, and Tom Heskes. "Optimal Training of Mean Variance Estimation Neural Networks."

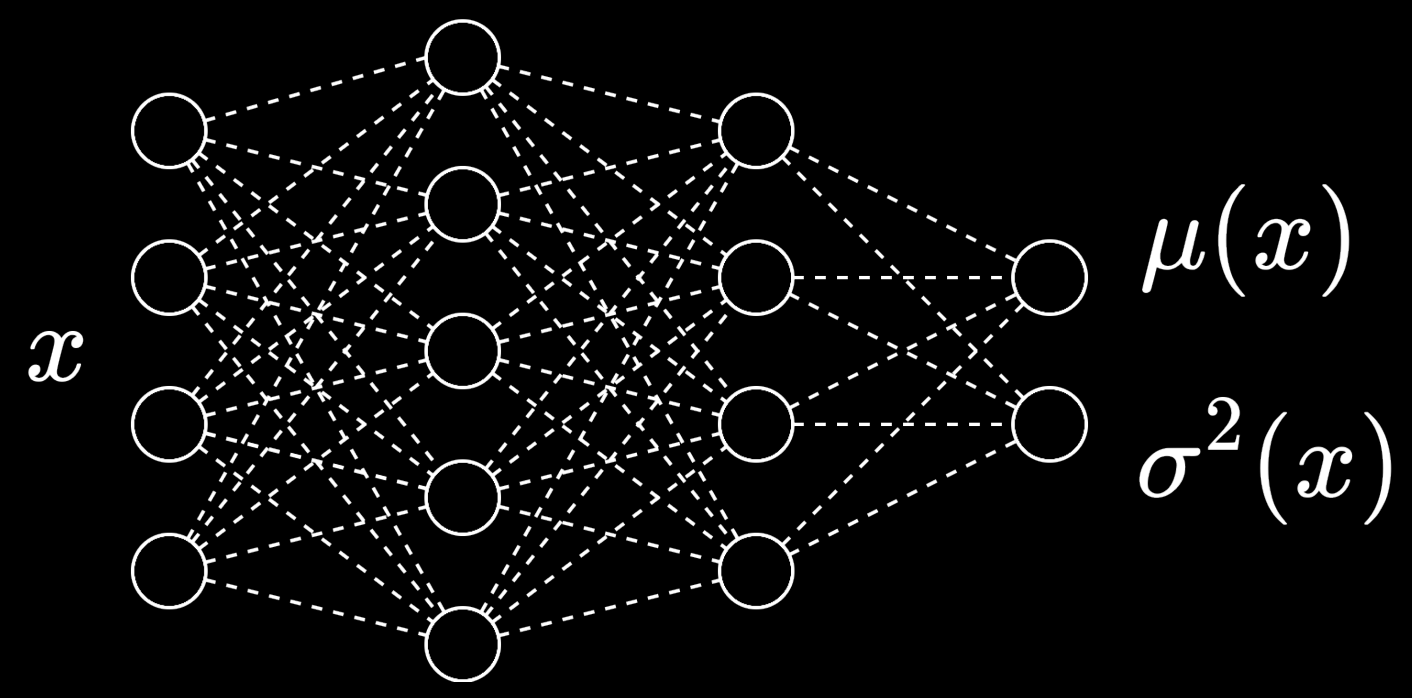
Sluijterman, Laurens, Eric Cator, and Tom Heskes. "Optimal Training of Mean Variance Estimation Neural Networks."



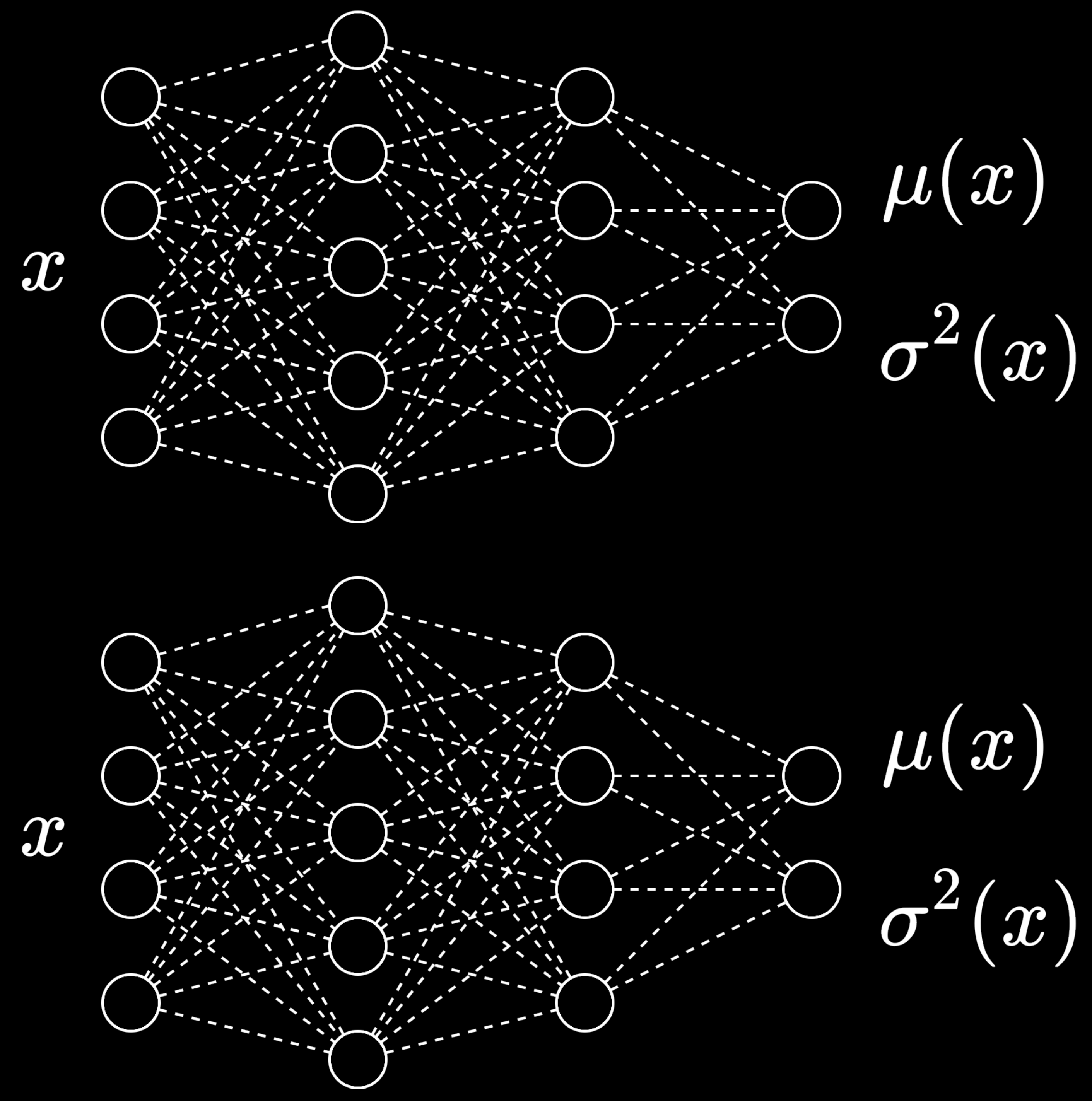
sin(x)
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
y
0
1
2
3
4
5
6
x

Deep Ensembles

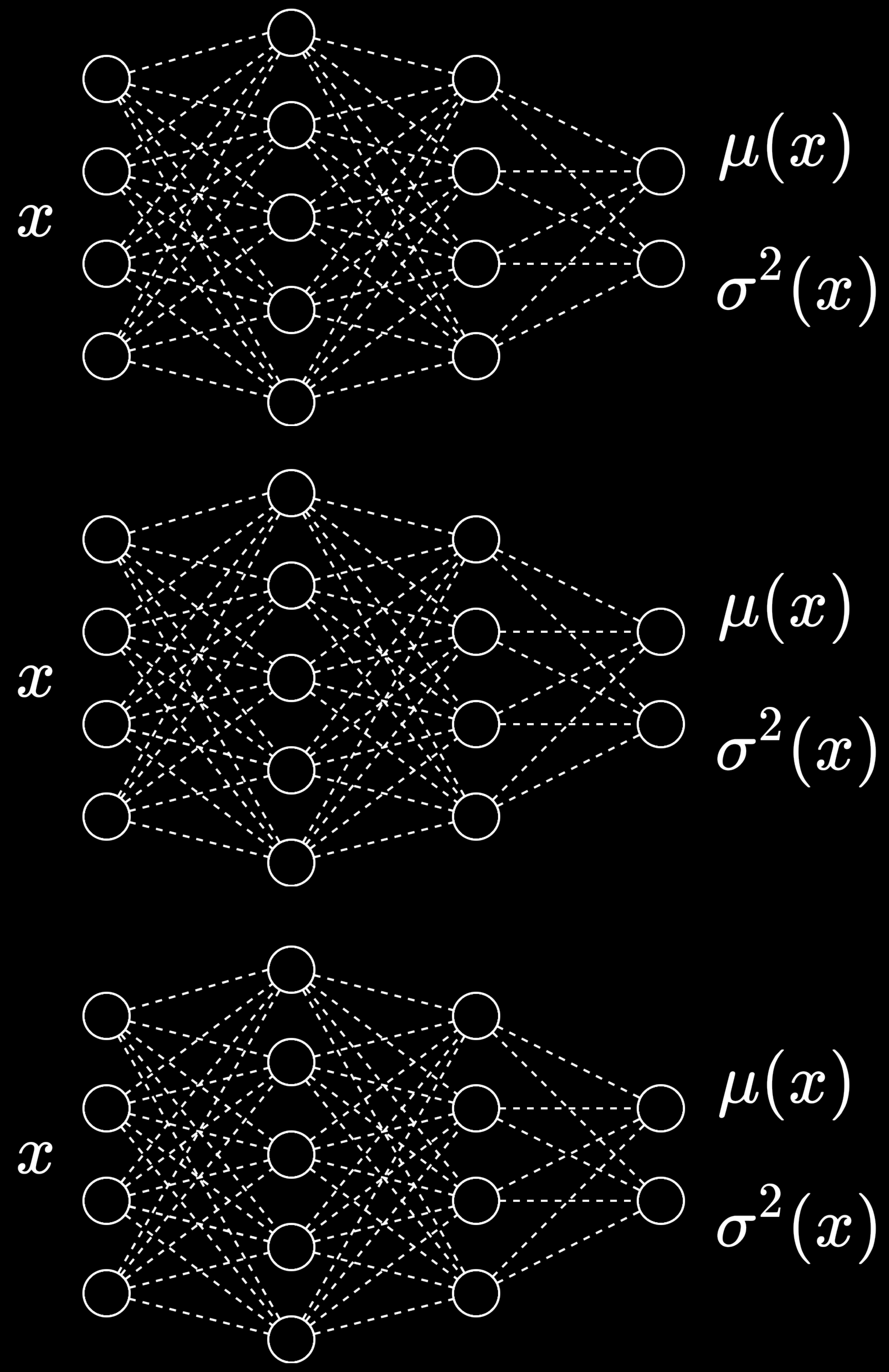
Deep Ensembles



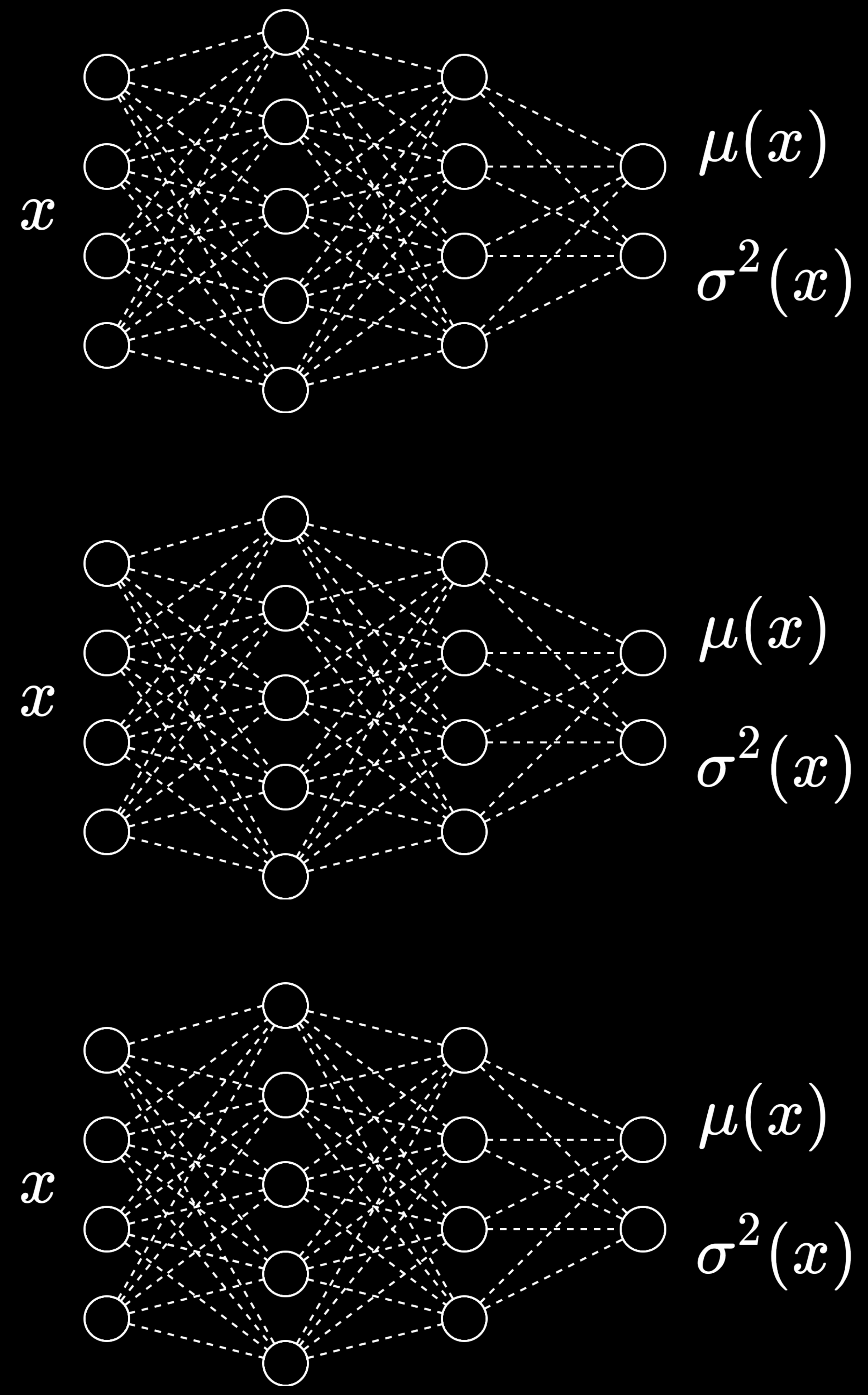
Deep Ensembles



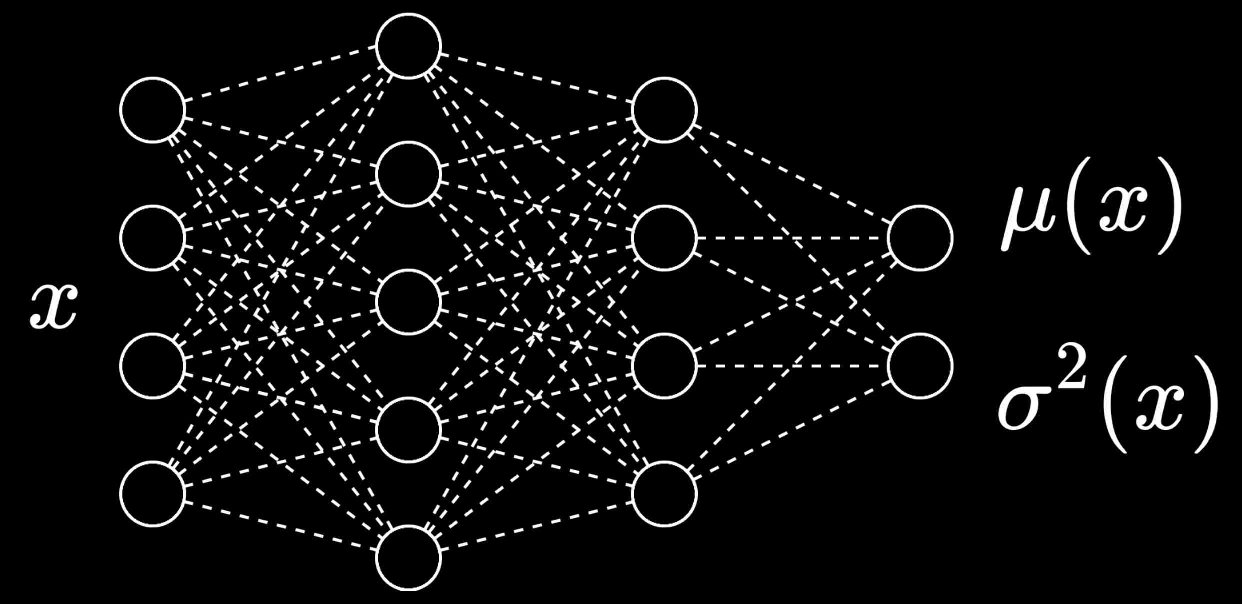
Deep Ensembles



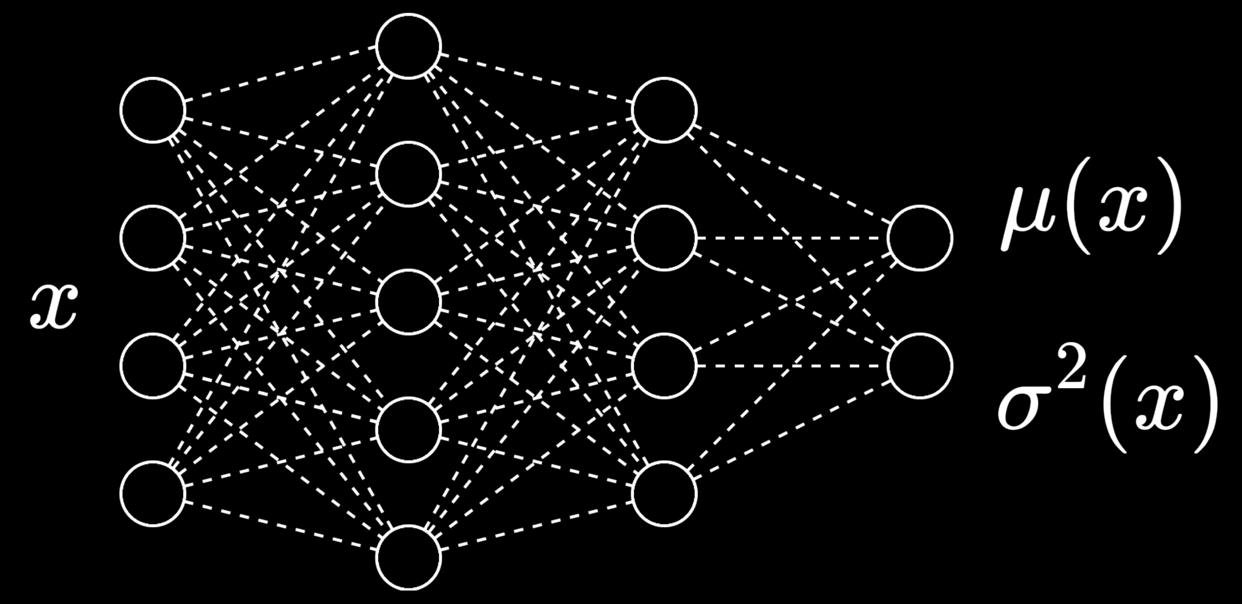
Parametric bootstrap



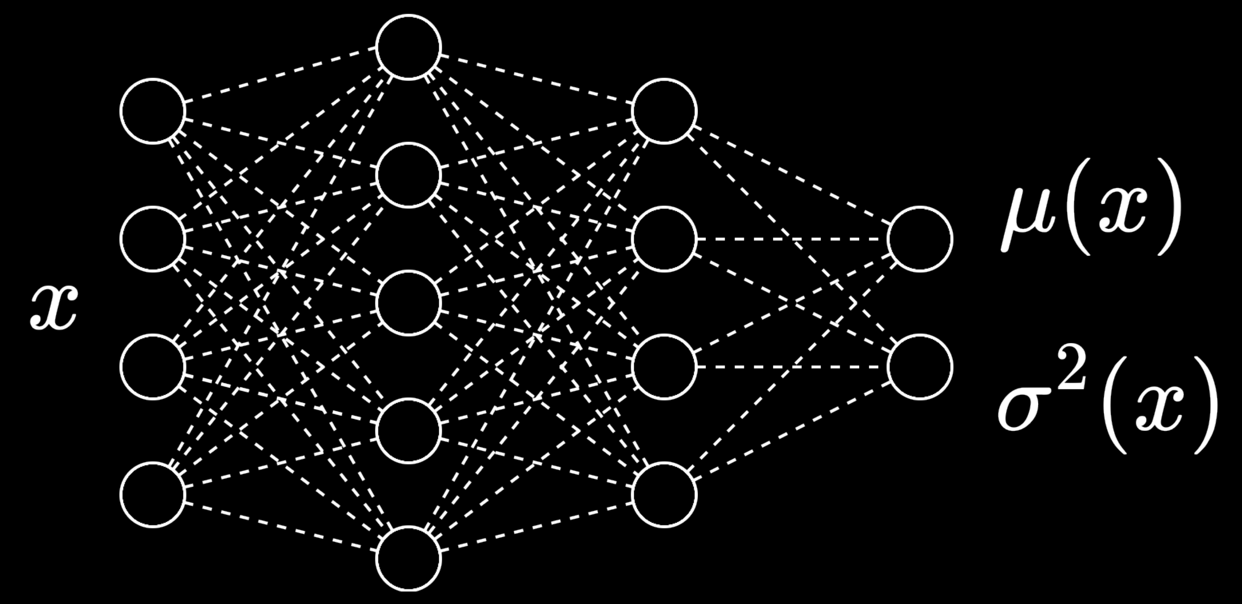
Parametric bootstrap



New data set

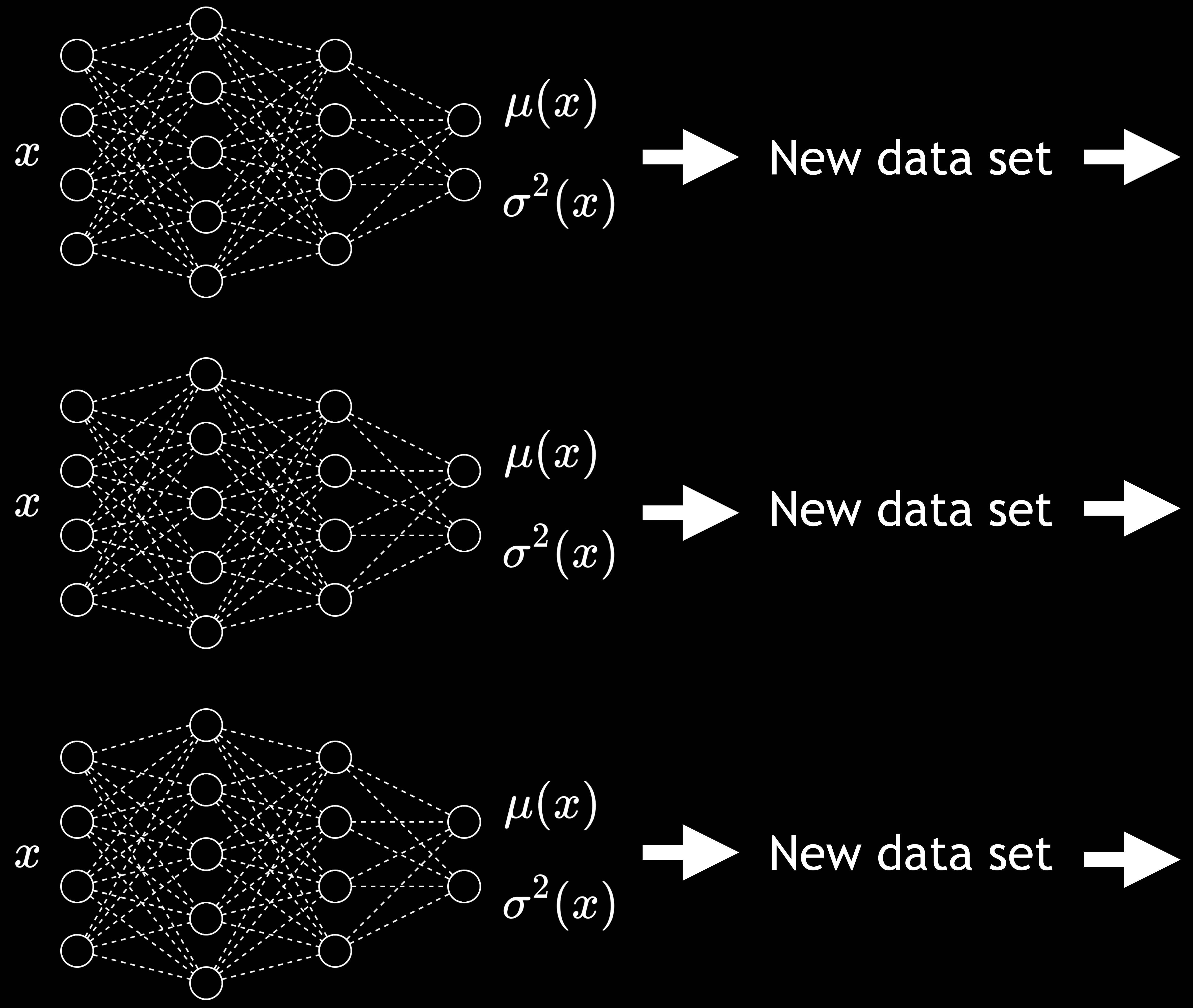


New data set

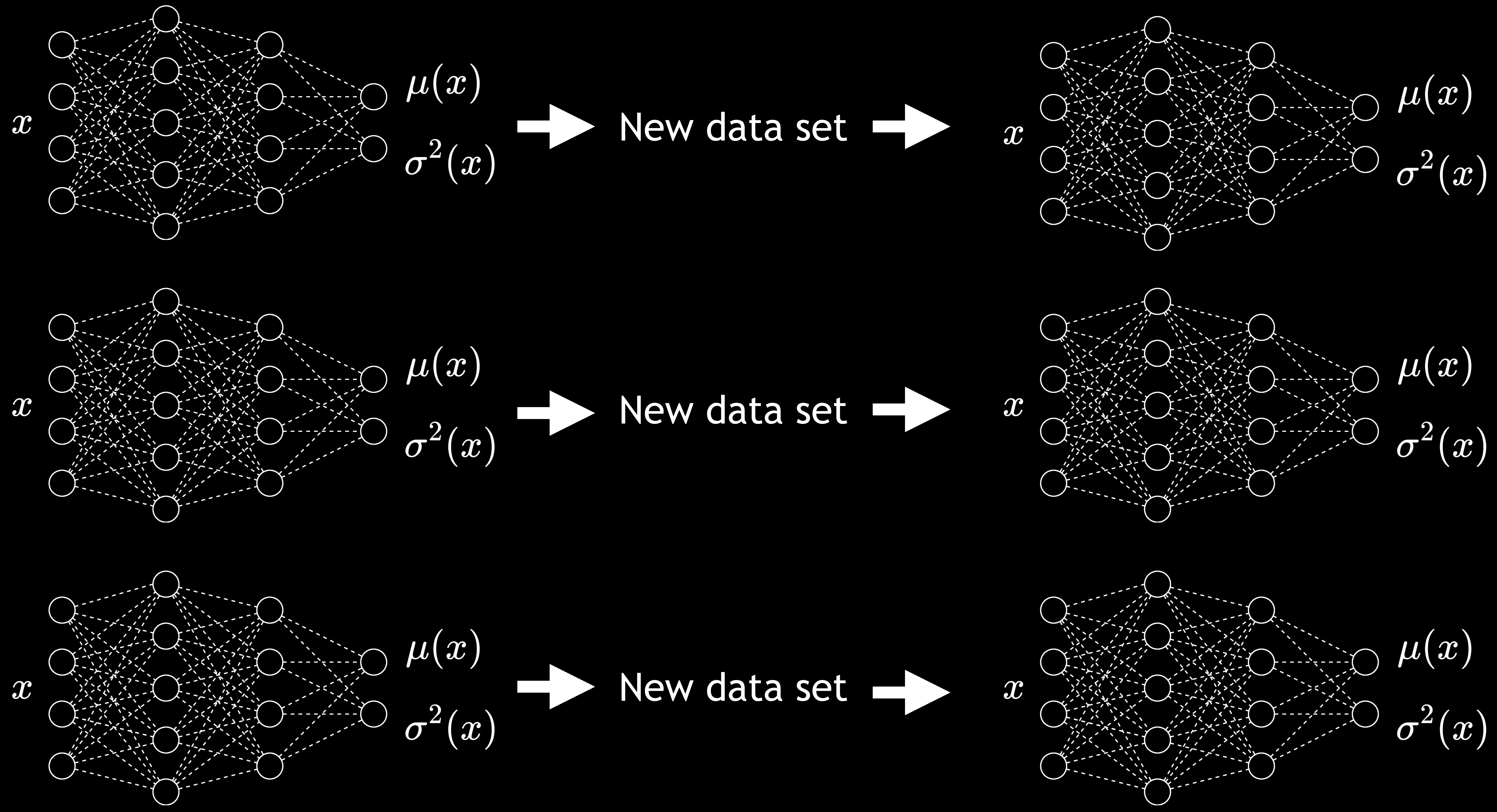


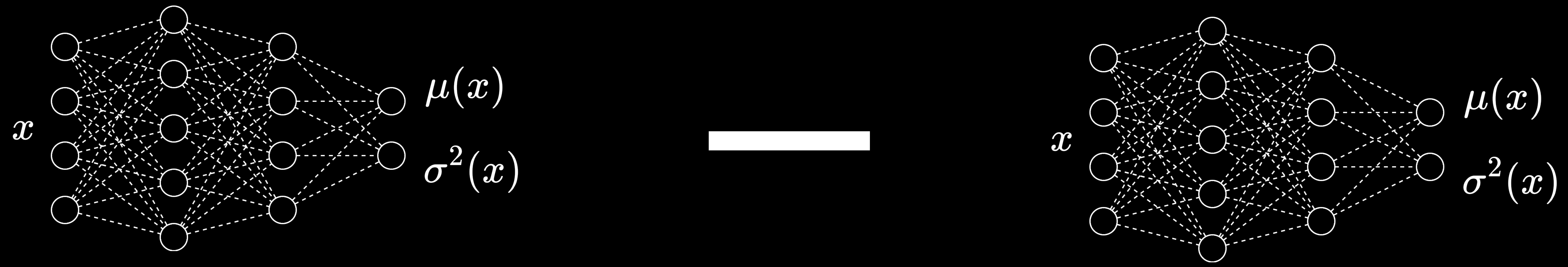
New data set

Parametric bootstrap



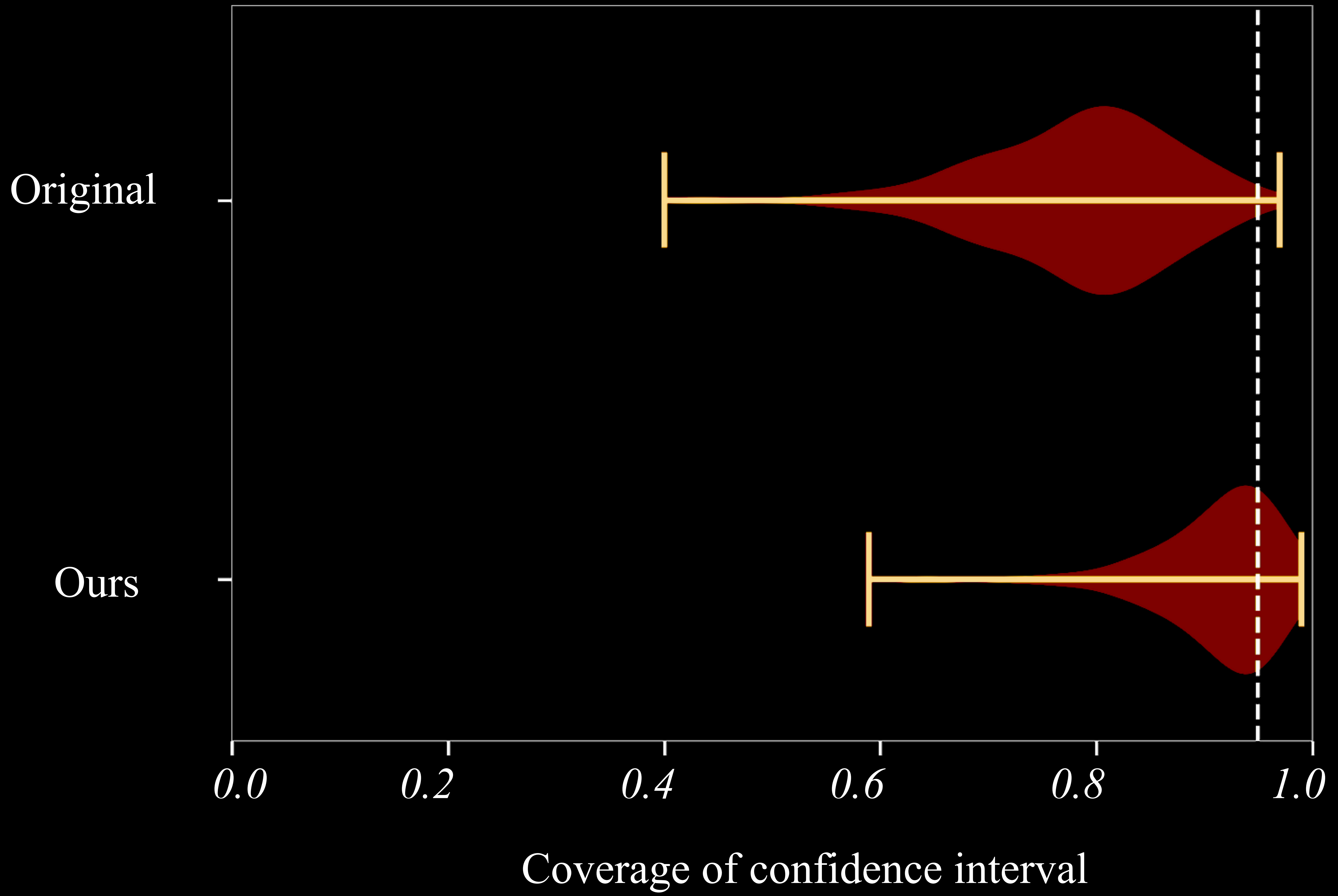
Parametric bootstrap





Evaluate the difference before and after retraining

Improved coverage



Sluijterman, Laurens, Eric Cator, and Tom Heskes. "Confident neural network regression with bootstrapped deep ensembles."

Bayesian

$$p(\theta \mid \mathscr{D}) = \frac{p(\mathscr{D} \mid \theta)p(\theta)}{p(\mathscr{D})}$$

Variational Inference

Variational Inference

1. Approximate $p(\theta \mid \mathscr{D})$ with $q_m(\theta)$

Variational Inference

1. Approximate $p(\theta \mid \mathscr{D})$ with $q_m(\theta)$
2. Minimize $\mathrm{KL}(q_m(\theta) \mid\mid p(\theta \mid \mathscr{D}))$

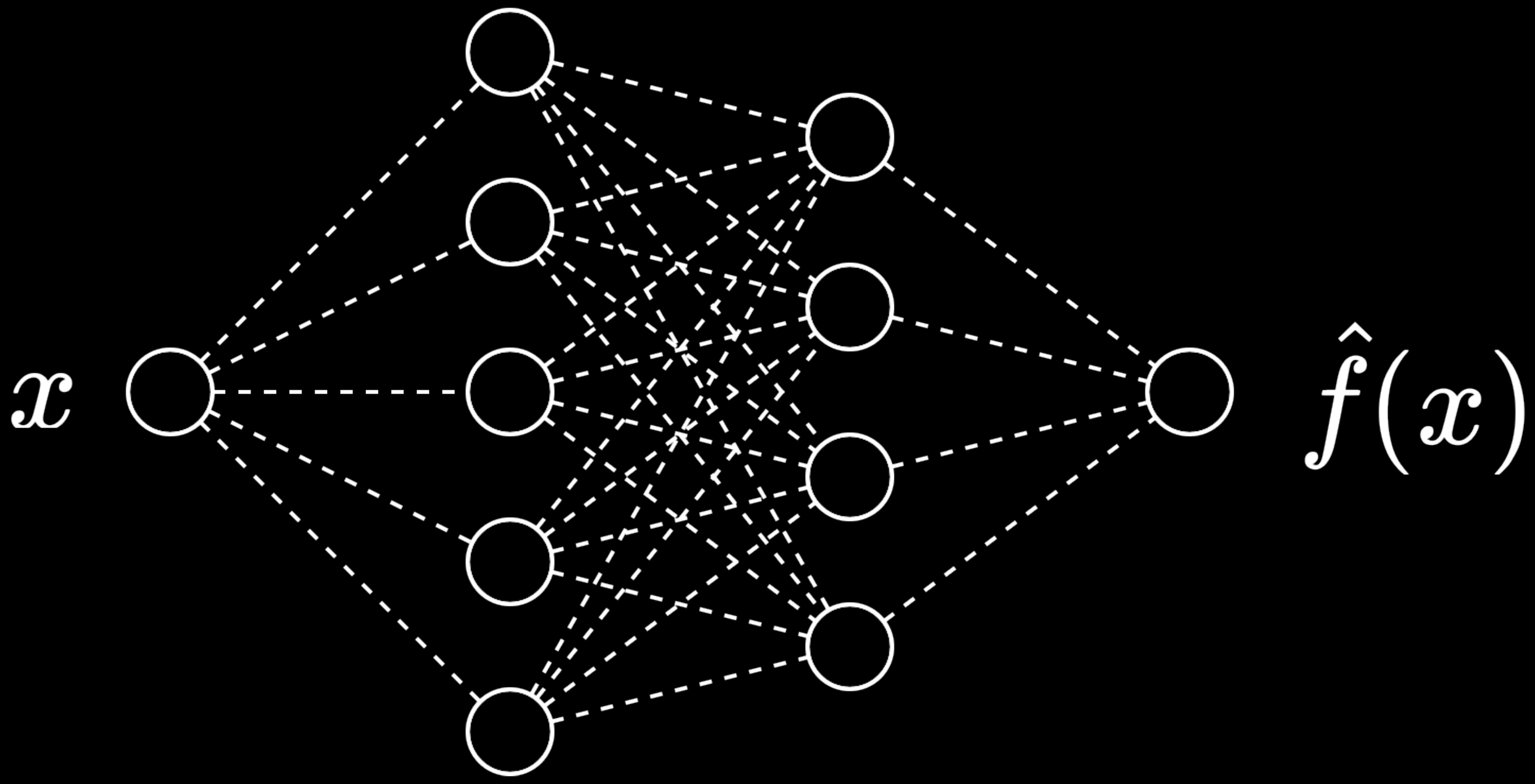
Variational Inference

1. Approximate $p(\theta \mid \mathscr{D})$ with $q_m(\theta)$
2. Minimize $\mathrm{KL}(q_m(\theta) \mid\mid p(\theta \mid \mathscr{D}))$

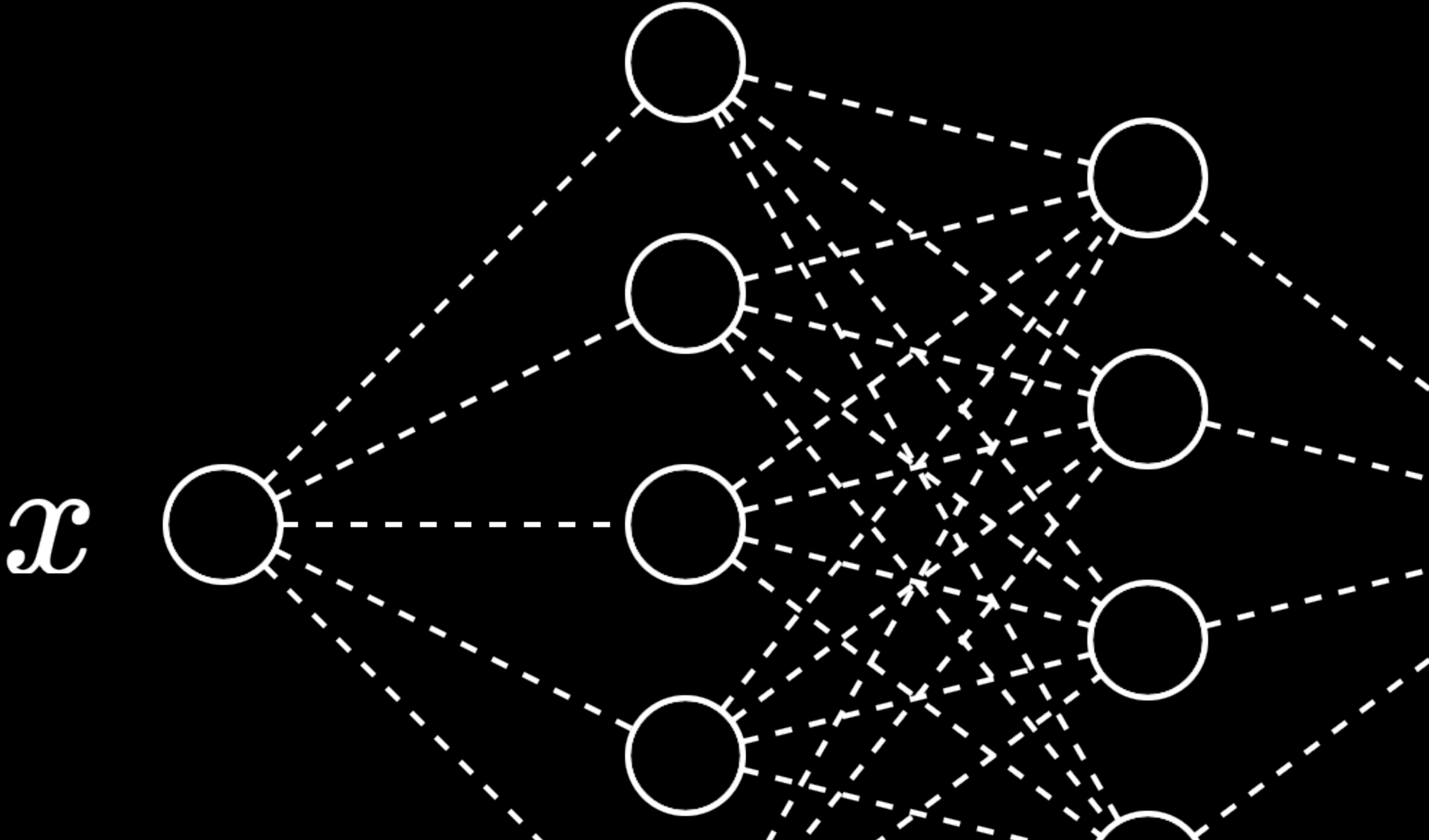
Variational Inference

1. Approximate $p(\theta \mid \mathscr{D})$ with $q_m(\theta)$
2. Minimize $\mathrm{KL}(q_m(\theta) \mid\mid p(\theta \mid \mathscr{D}))$

Dropout

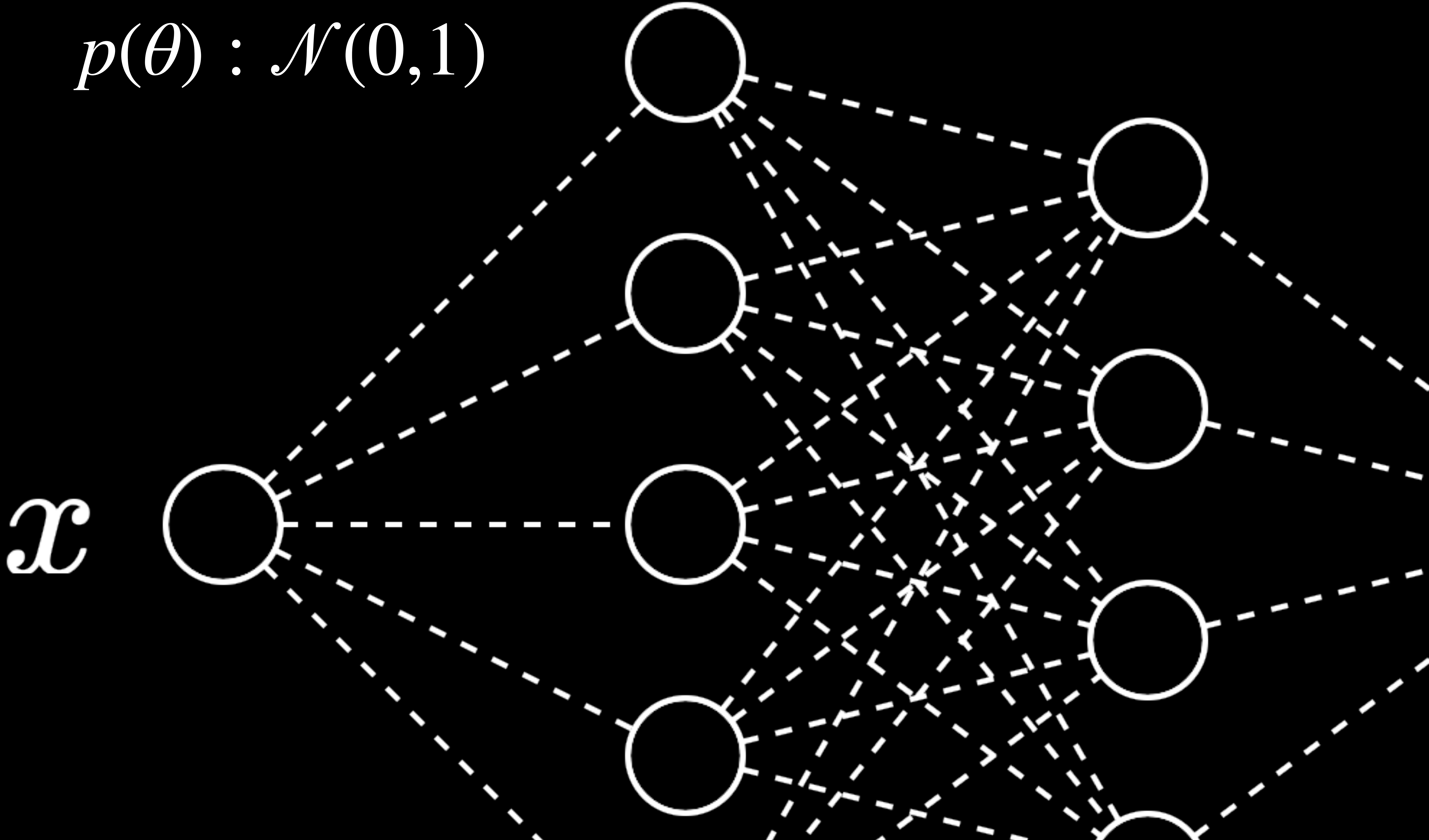


Dropout



Dropout

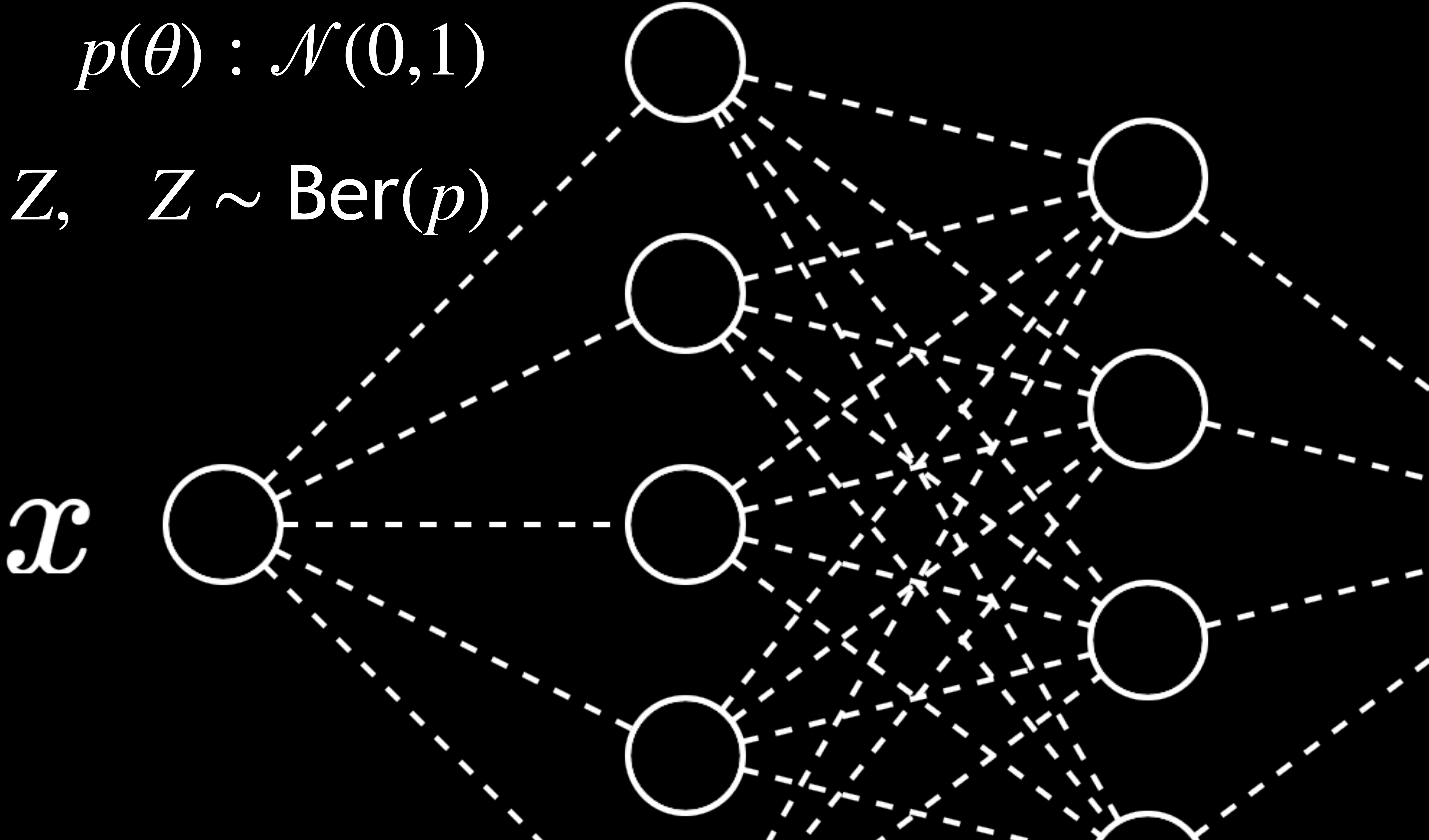
$$p(\theta) : \mathcal{N}(0,1)$$



Dropout

$$p(\theta) : \mathcal{N}(0,1)$$

$$q_m(\theta) : m \cdot Z, \quad Z \sim \mathrm{Ber}(p)$$



Interpretation

Interpretation

1. What does a high-dimensional prior over weights mean?

Interpretation

1. What does a high-dimensional prior over weights mean?
2. The prior is chosen in order to get the result.

Interpretation

1. What does a high-dimensional prior over weights mean?
2. The prior is chosen in order to get the result.
3. How good is the approximation?

Interpretation

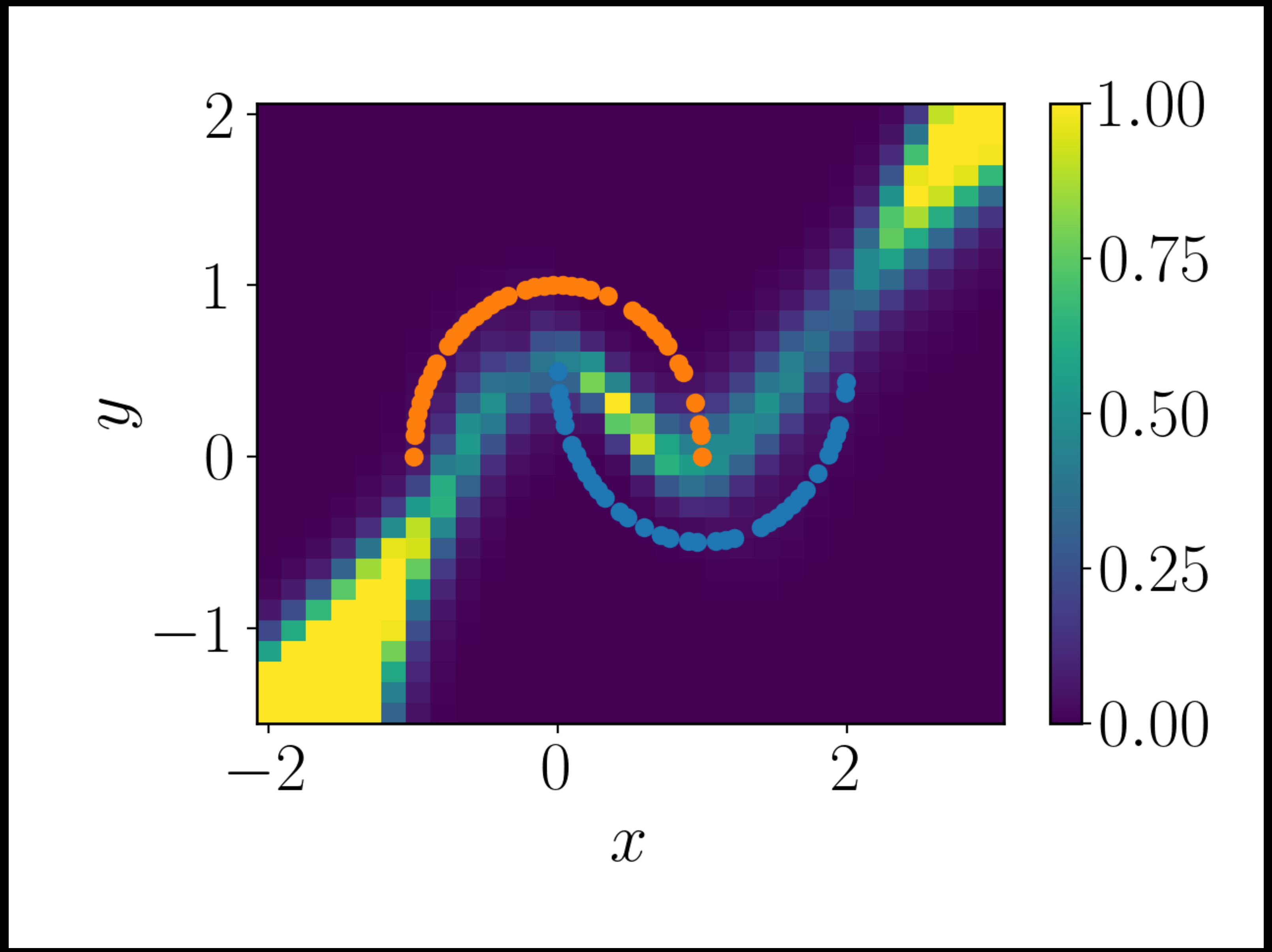
1. What does a high-dimensional prior over weights mean?
2. The prior is chosen in order to get the result.
3. How good is the approximation?
4. Random training

Interpretation

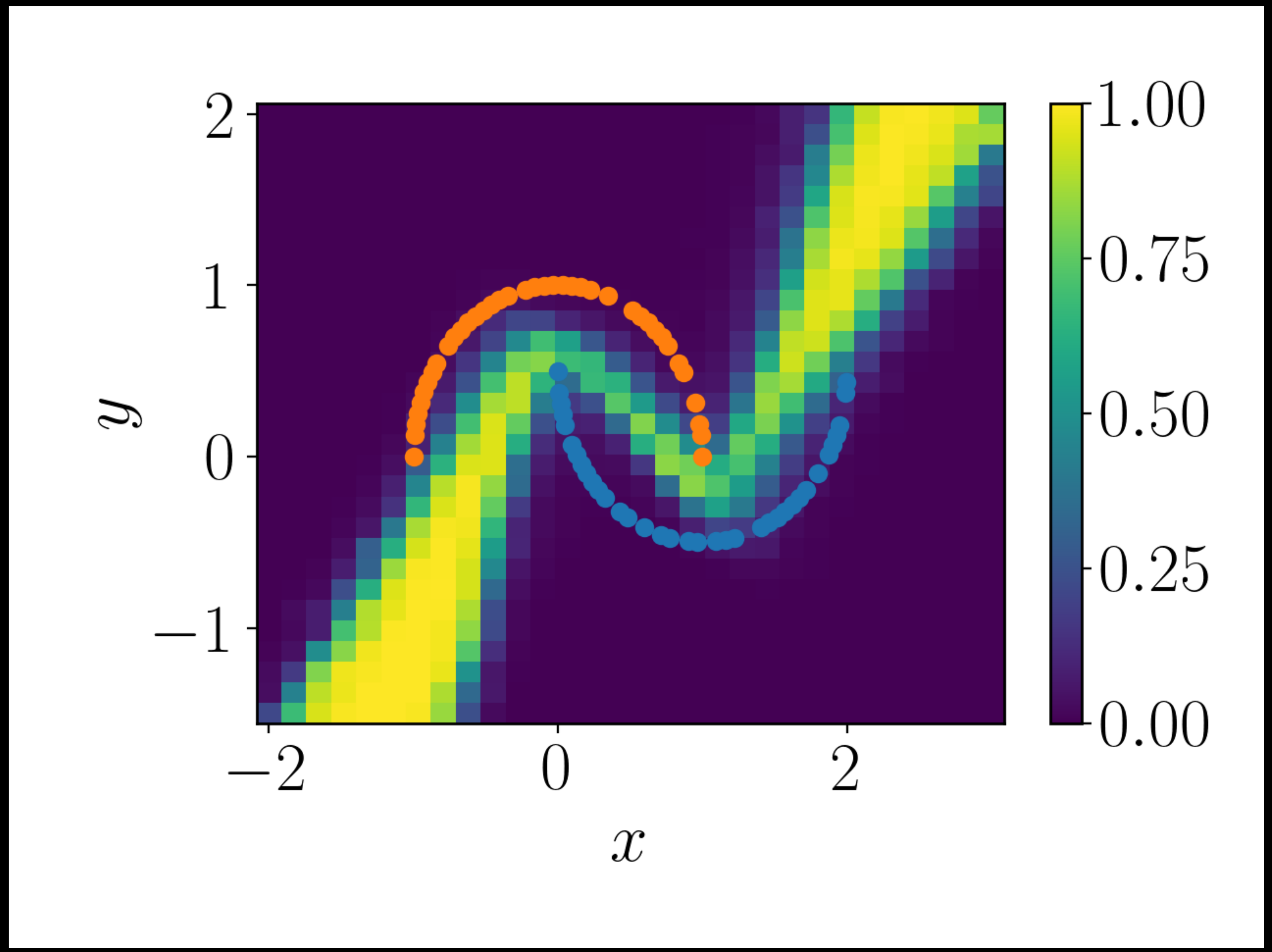
1. What does a high-dimensional prior over weights mean?
2. The prior is chosen in order to get the result.
3. How good is the approximation?
4. Random training

Out-of-distribution

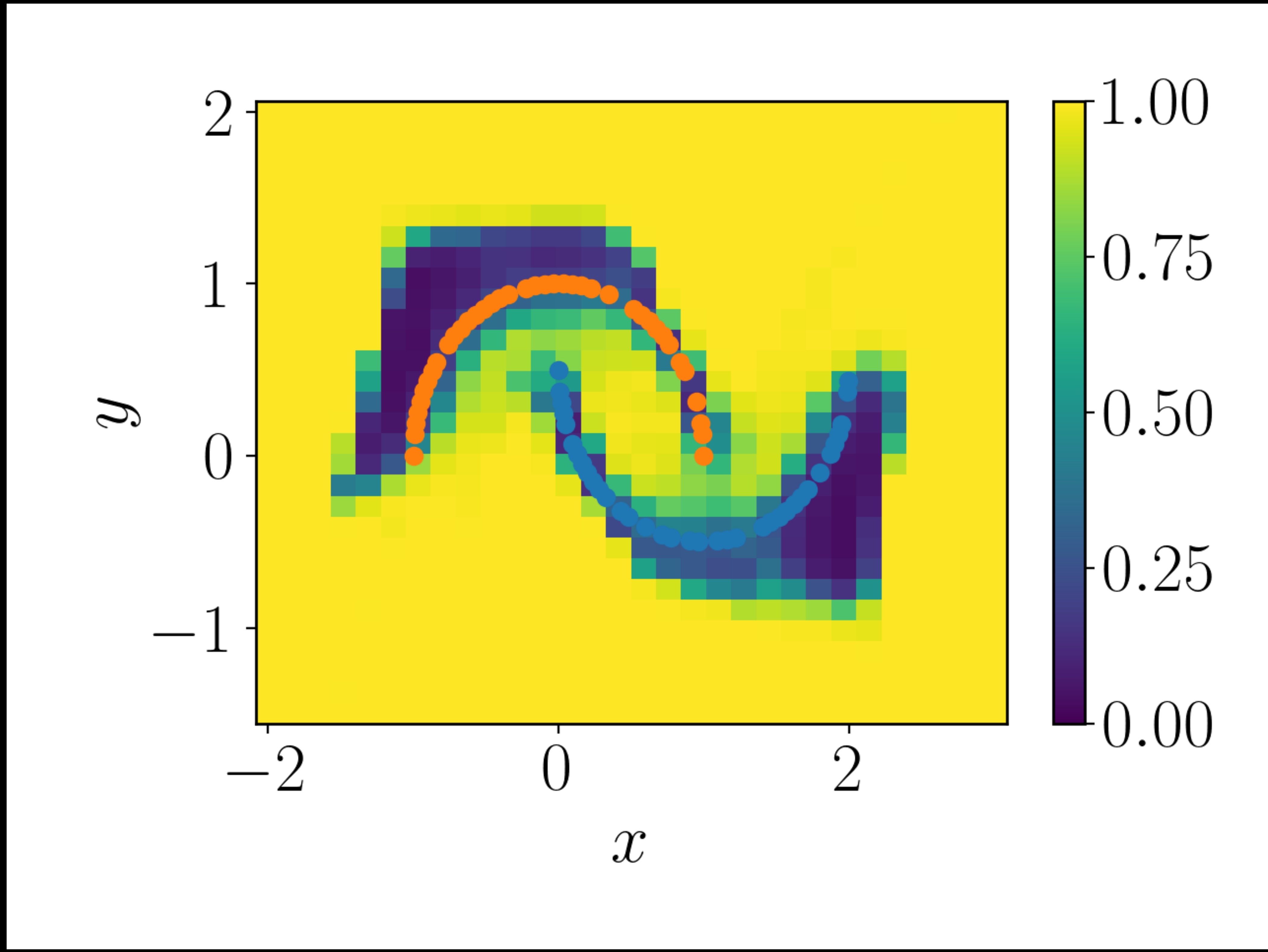
Ensemble



Dropout



Likelihood ratio



Sluijterman, Laurens, Eric Cator, and Tom Heskes. "Likelihood-ratio-based confidence intervals for neural networks."

Summary

Summary

1. All sources of uncertainty?

Summary

1. All sources of uncertainty?

Summary

1. All sources of uncertainty?

2. Is our model correct?

Summary

1. All sources of uncertainty?

2. Is our model correct?

Summary

1. All sources of uncertainty?

2. Is our model correct?

3. MVE network and Out-of-distribution

Summary

1. All sources of uncertainty?

2. Is our model correct?

3. MVE network and Out-of-distribution