E9 205 Machine Learning for Signal Processing

Neural Networks - Generalization

28-10-2019

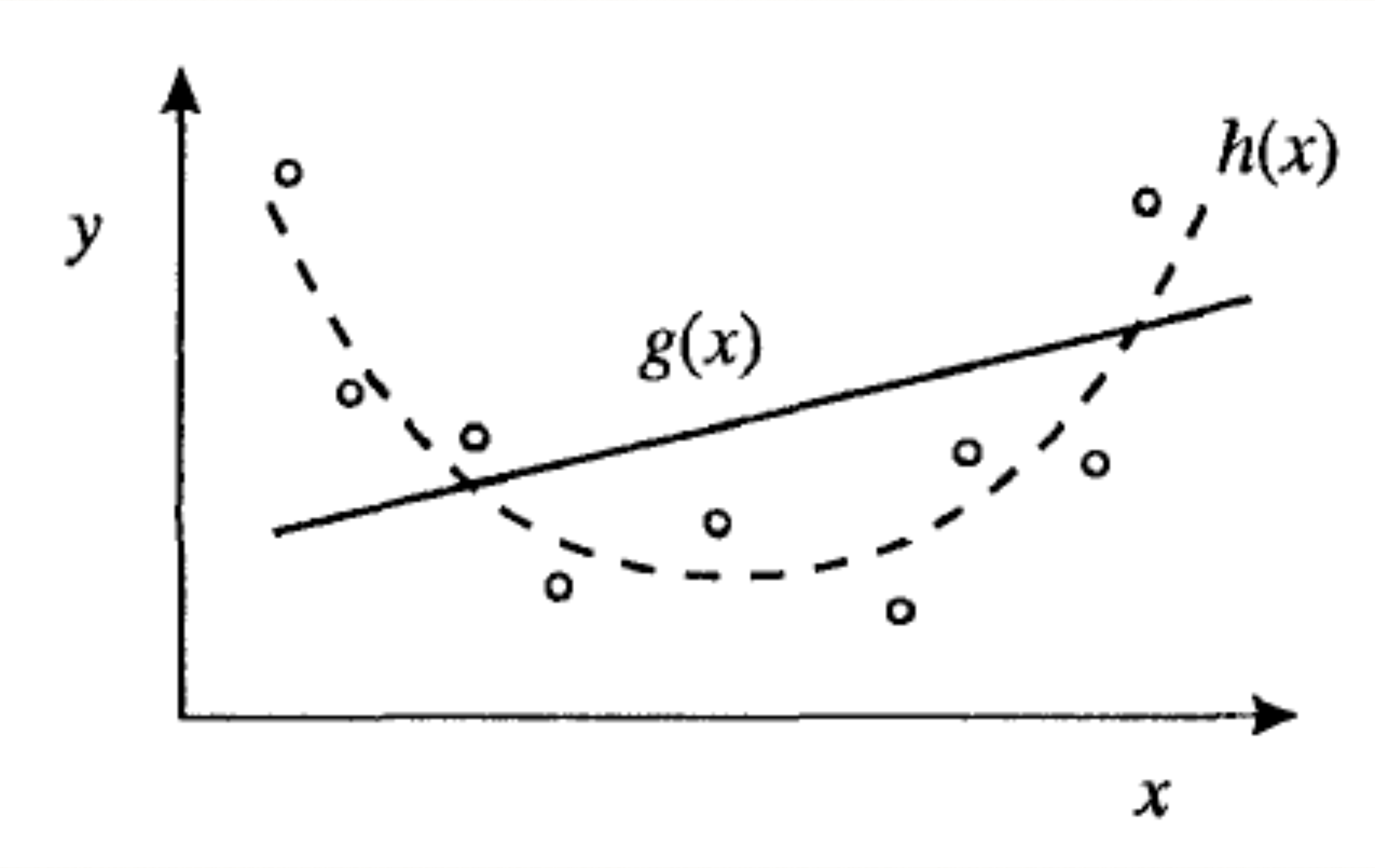
Instructor - Sriram Ganapathy (sriramg@iisc.ac.in.in)

Bias and Variance In Neural Network Training

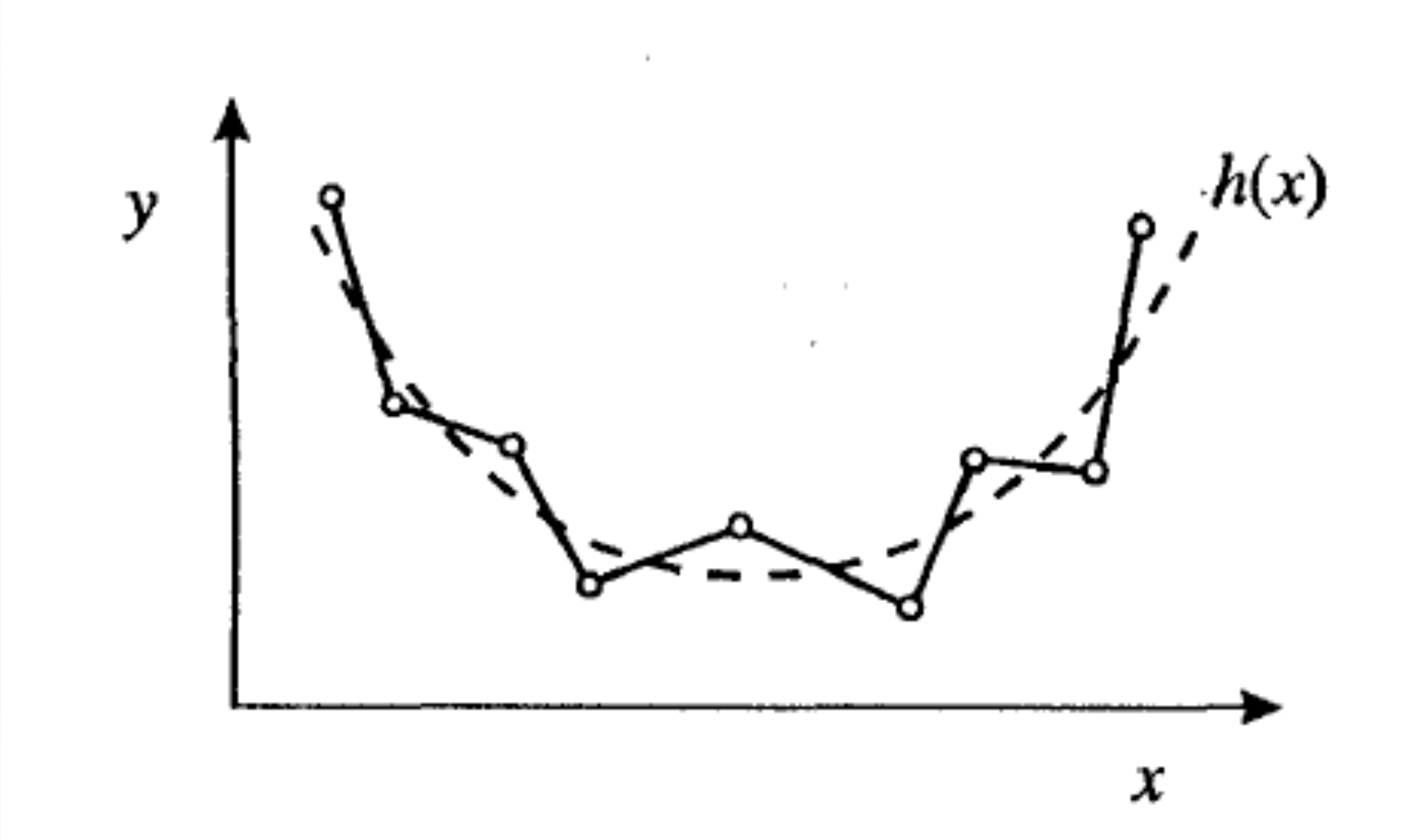
$$(\text{bias})^2 = \frac{1}{2}\int \{\mathcal{E}_D[y(\mathbf{x})] - \langle t|\mathbf{x}\rangle\}^2 p(\mathbf{x})\,d\mathbf{x}$$

$$\text{variance} = \frac{1}{2}\int \mathcal{E}_D[\{y(\mathbf{x}) - \mathcal{E}_D[y(\mathbf{x})]\}^2]p(\mathbf{x})\,d\mathbf{x}.$$

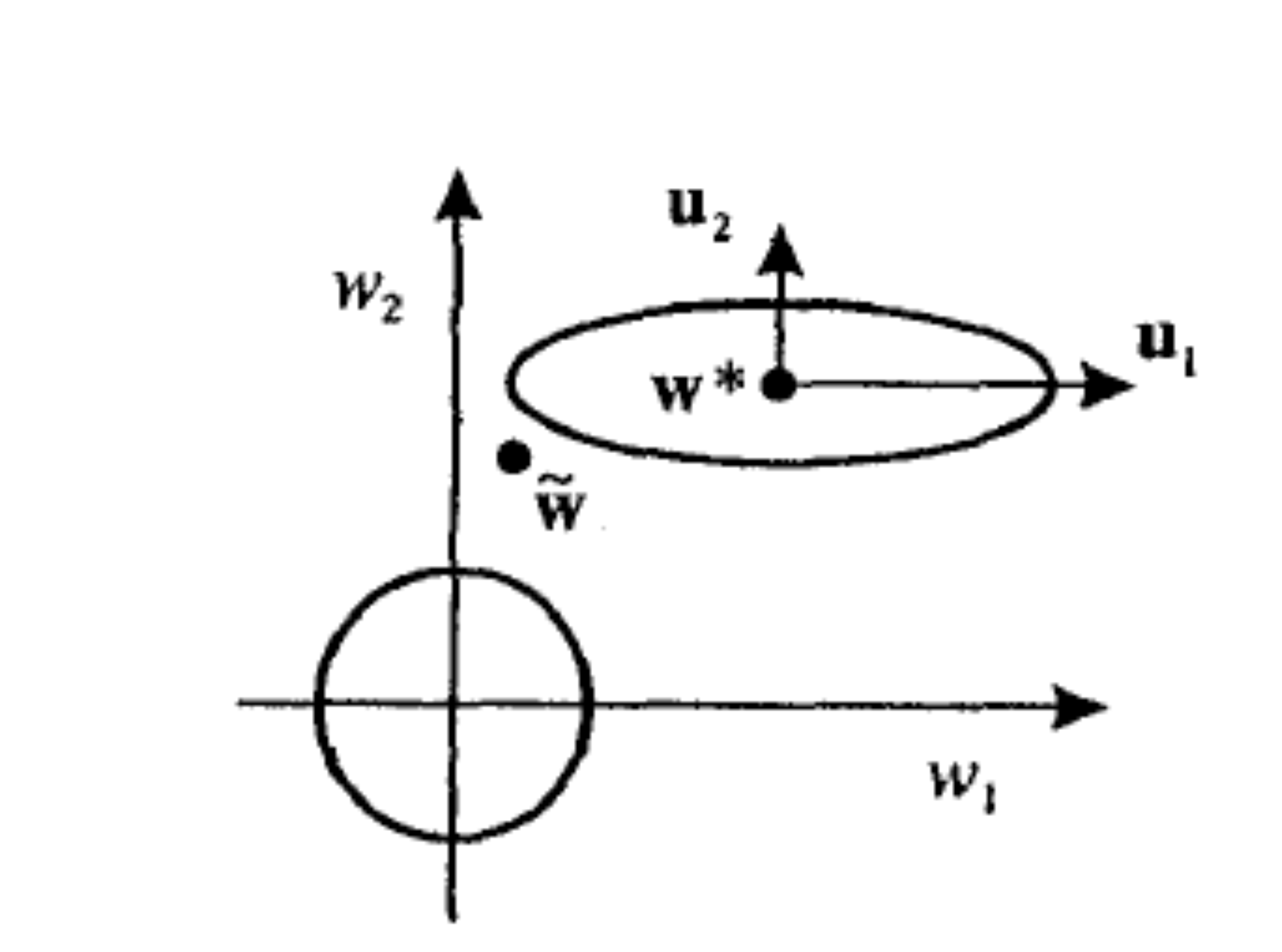
Underfit



Overfit

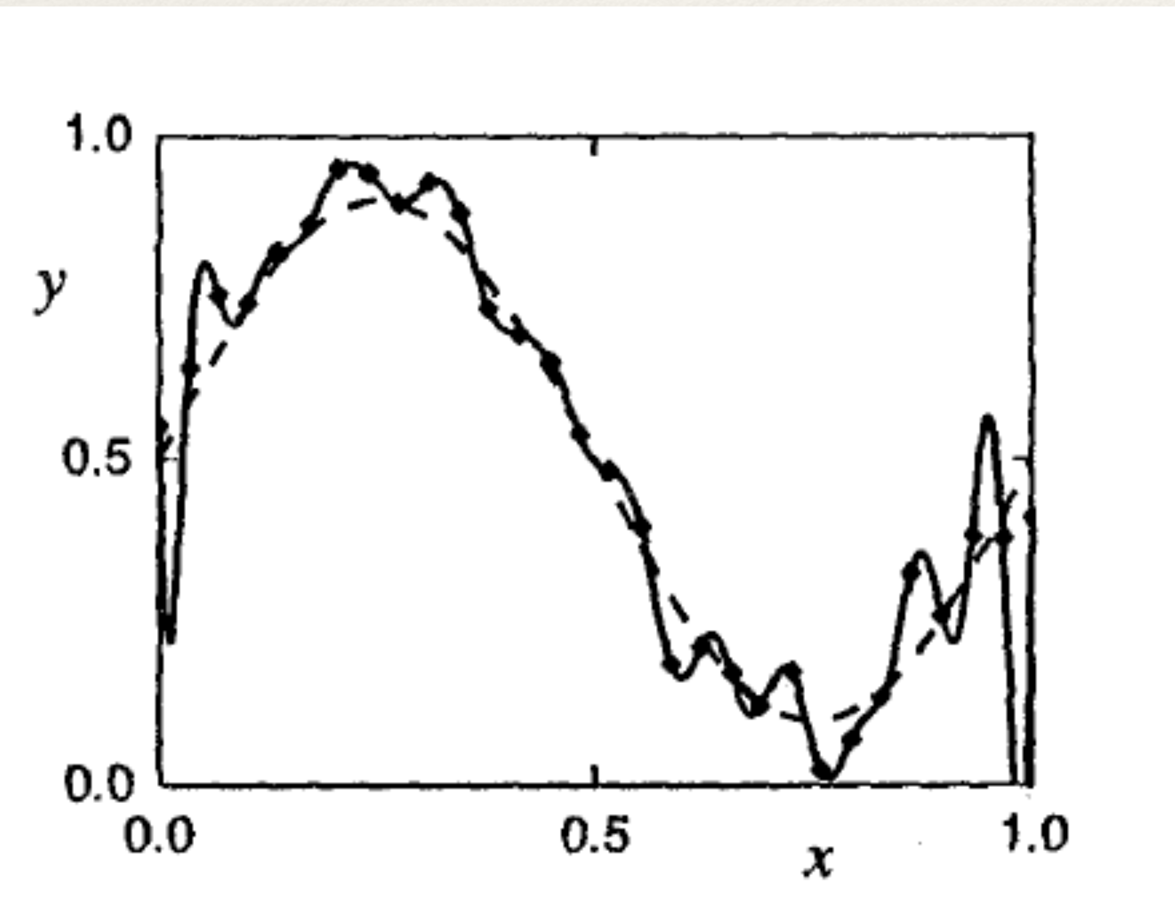


Weight Decay Based Regularization

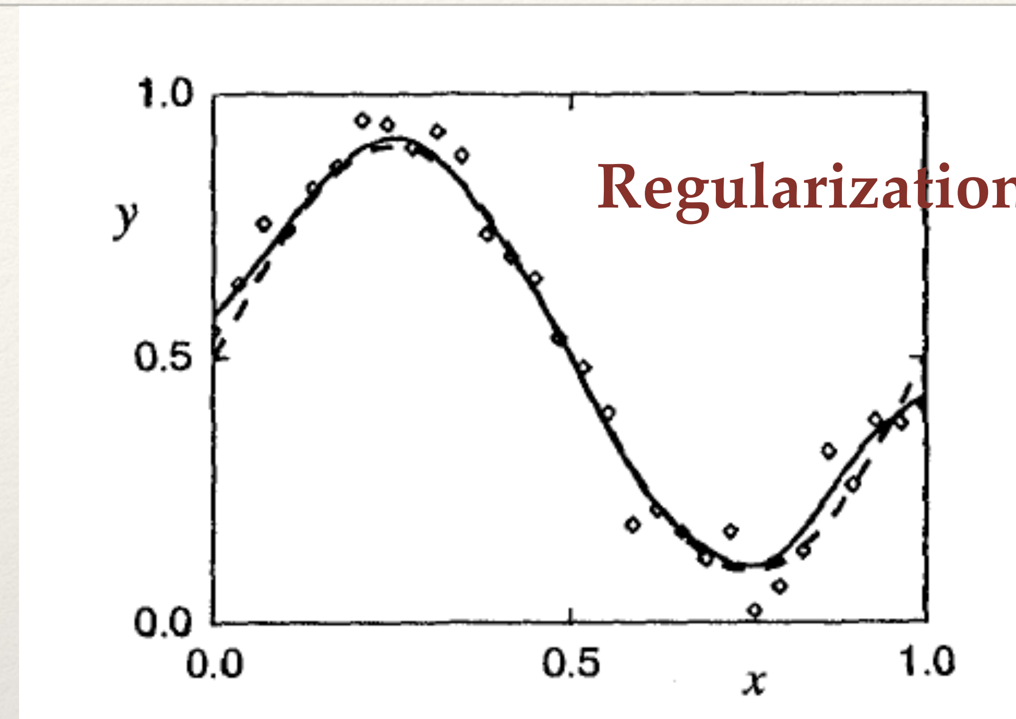


Weight Decay Regularization

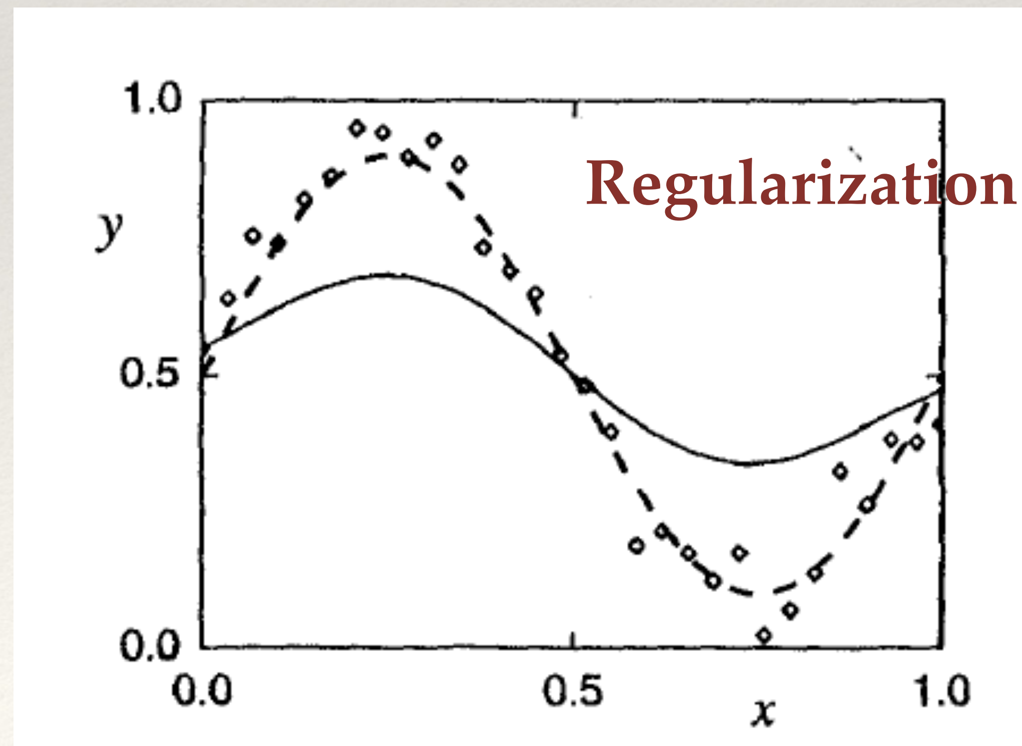
Regularization = 0



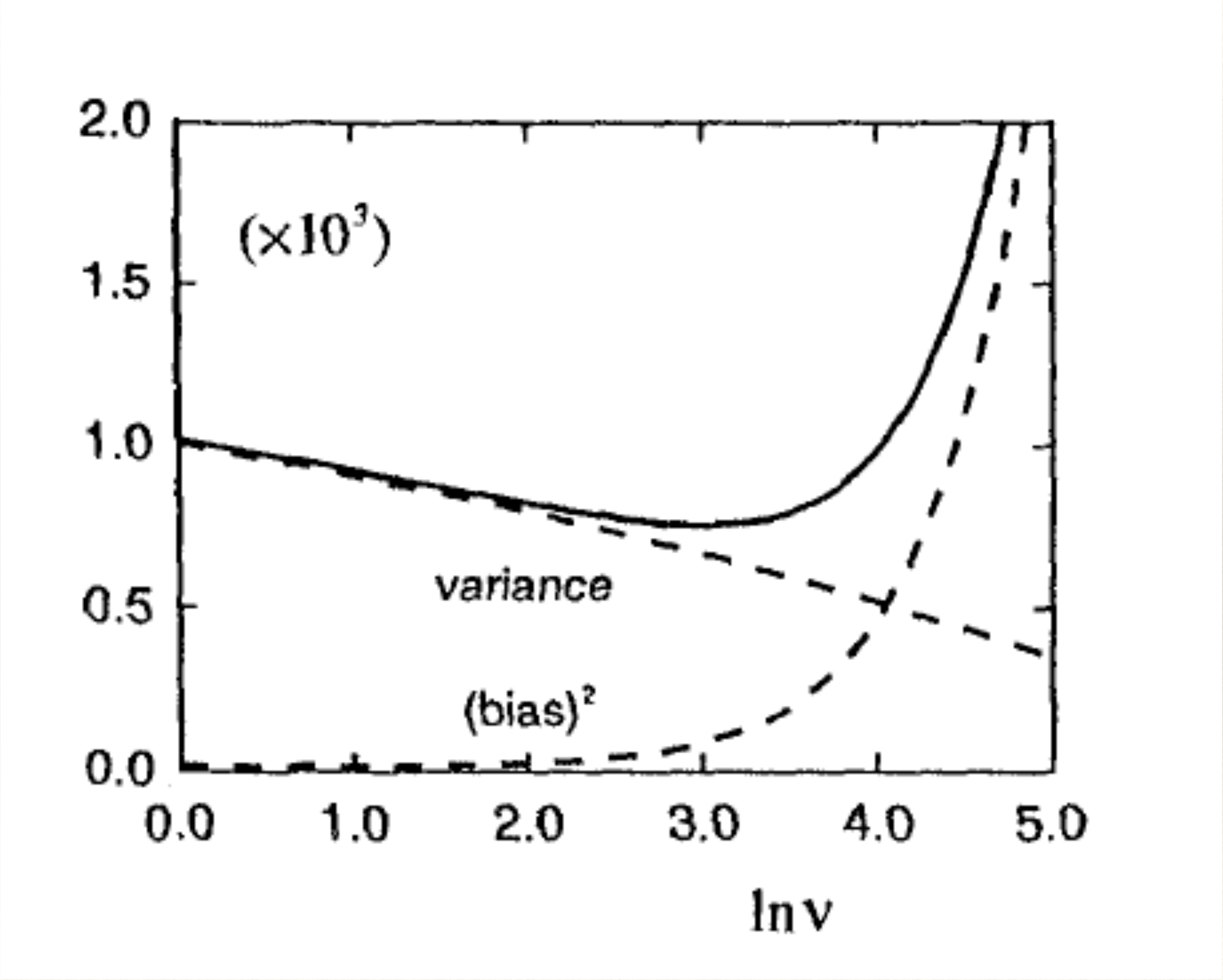
Regularization = 40



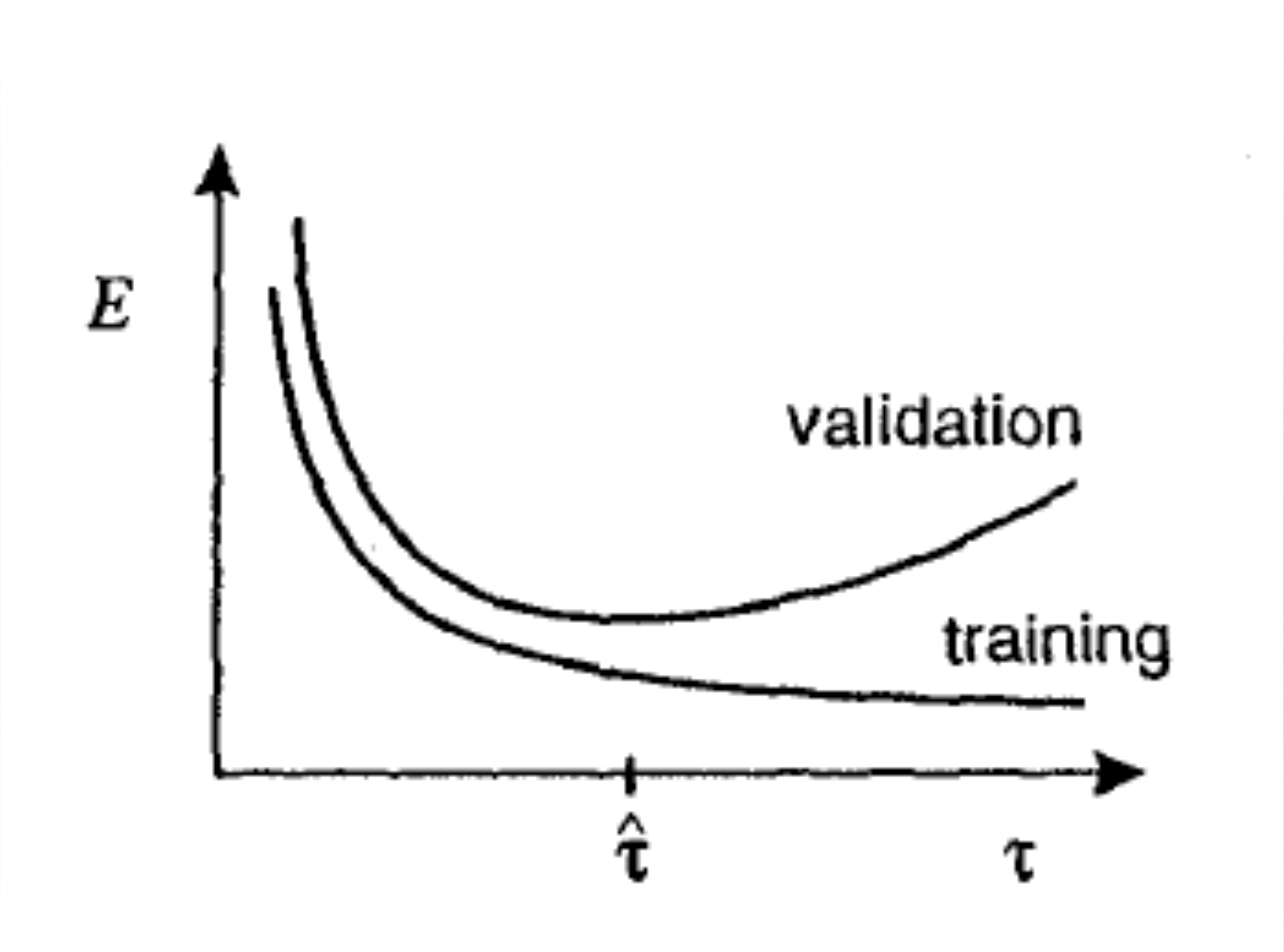
Regularization = 4000



Regularization Effect on Learning



Early Stopping



Most Popular in Practice

Neural Networks - Summary

- Details of Architecture
- Computation of gradient using back propagation.
- Error function and output layer activation
  - Neural networks estimate posterior probabilities
- Learning in Neural networks
  - Gradient descent - Properties
- Generalization of Neural Networks

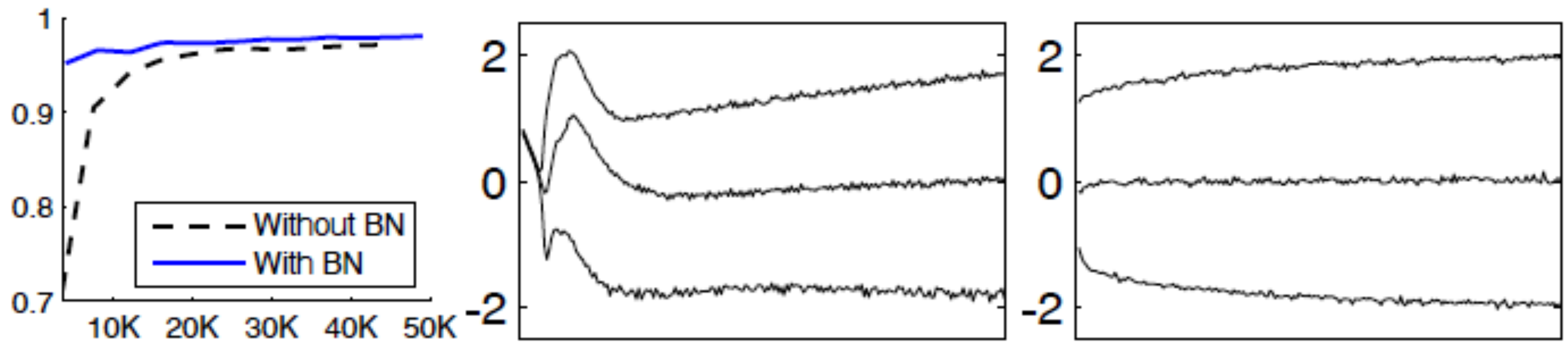
Batch Normalization

Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift

Sergey Ioffe
Google Inc., *sioffe@google.com*

Christian Szegedy
Google Inc., *szegedy@google.com*

Effect of Batch Normalization



(a) (b) Without BN (c) With BN

Figure 1: (a) *The test accuracy of the MNIST network trained with and without Batch Normalization, vs. the number of training steps. Batch Normalization helps the network train faster and achieve higher accuracy.* (b, c) *The evolution of input distributions to a typical sigmoid, over the course of training, shown as* $\{15, 50, 85\}$*th percentiles. Batch Normalization makes the distribution more stable and reduces the internal covariate shift.*

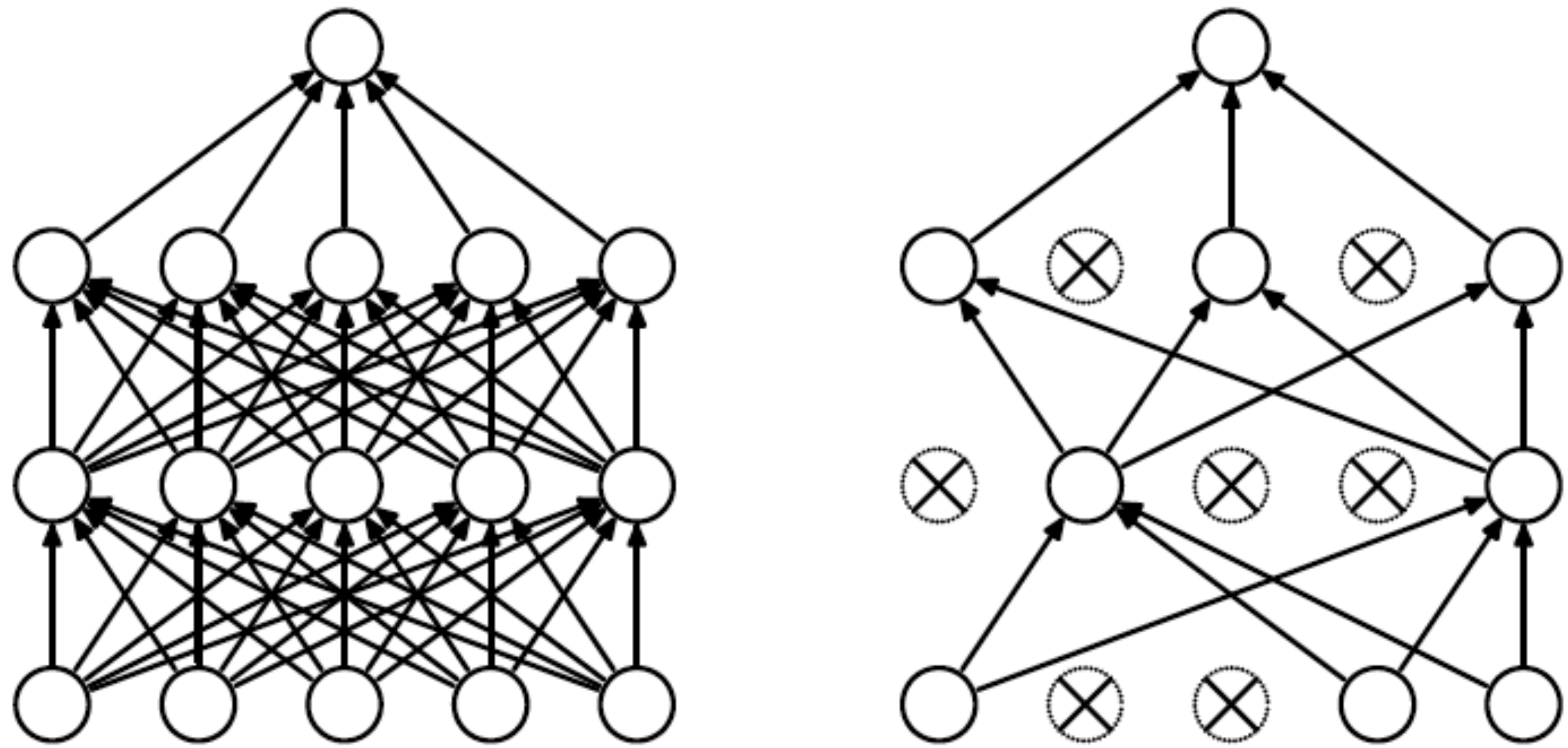
Dropout Strategy in Neural Network Training

Dropout: A Simple Way to Prevent Neural Networks from Overfitting

Nitish Srivastava NITISH@CS.TORONTO.EDU
Geoffrey Hinton HINTON@CS.TORONTO.EDU
Alex Krizhevsky KRIZ@CS.TORONTO.EDU
Ilya Sutskever ILYA@CS.TORONTO.EDU
Ruslan Salakhutdinov RSALAKHU@CS.TORONTO.EDU
Department of Computer Science
University of Toronto
10 Kings College Road, Rm 3302
Toronto, Ontario, M5S 3G4, Canada.

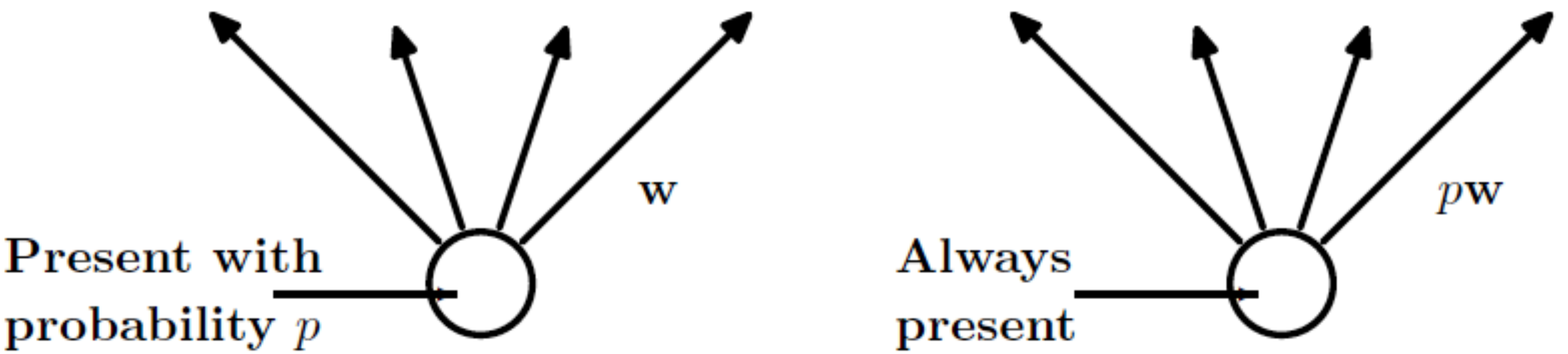
Editor: Yoshua Bengio

Dropouts in Neural Networks



(a) Standard Neural Net (b) After applying dropout.

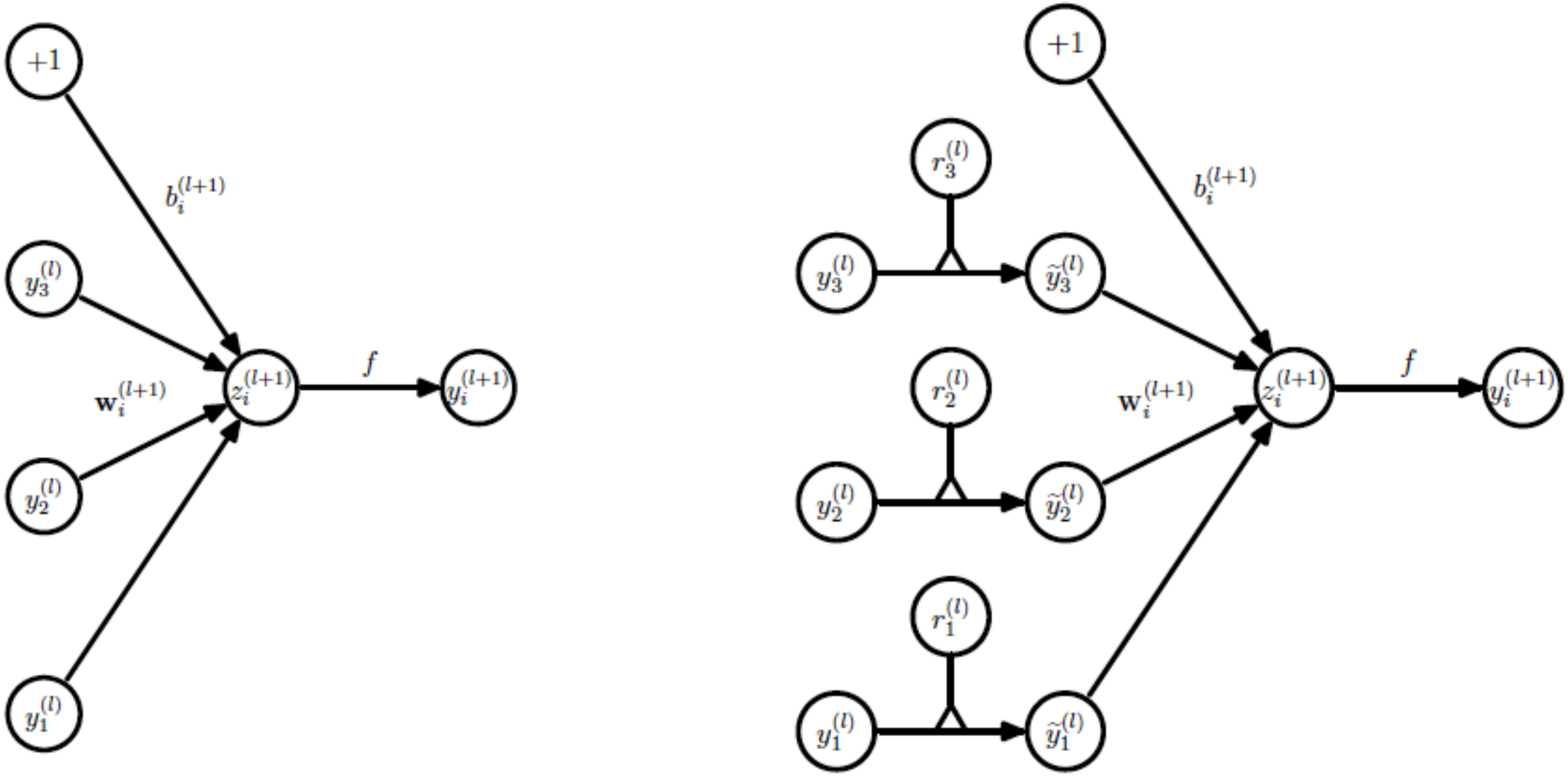
Dropout in Training and Test



(a) At training time

(b) At test time

Dropout Application



(a) Standard network

(b) Dropout network

Figure 3: Comparison of the basic operations of a standard and dropout network.

Effect of Dropouts

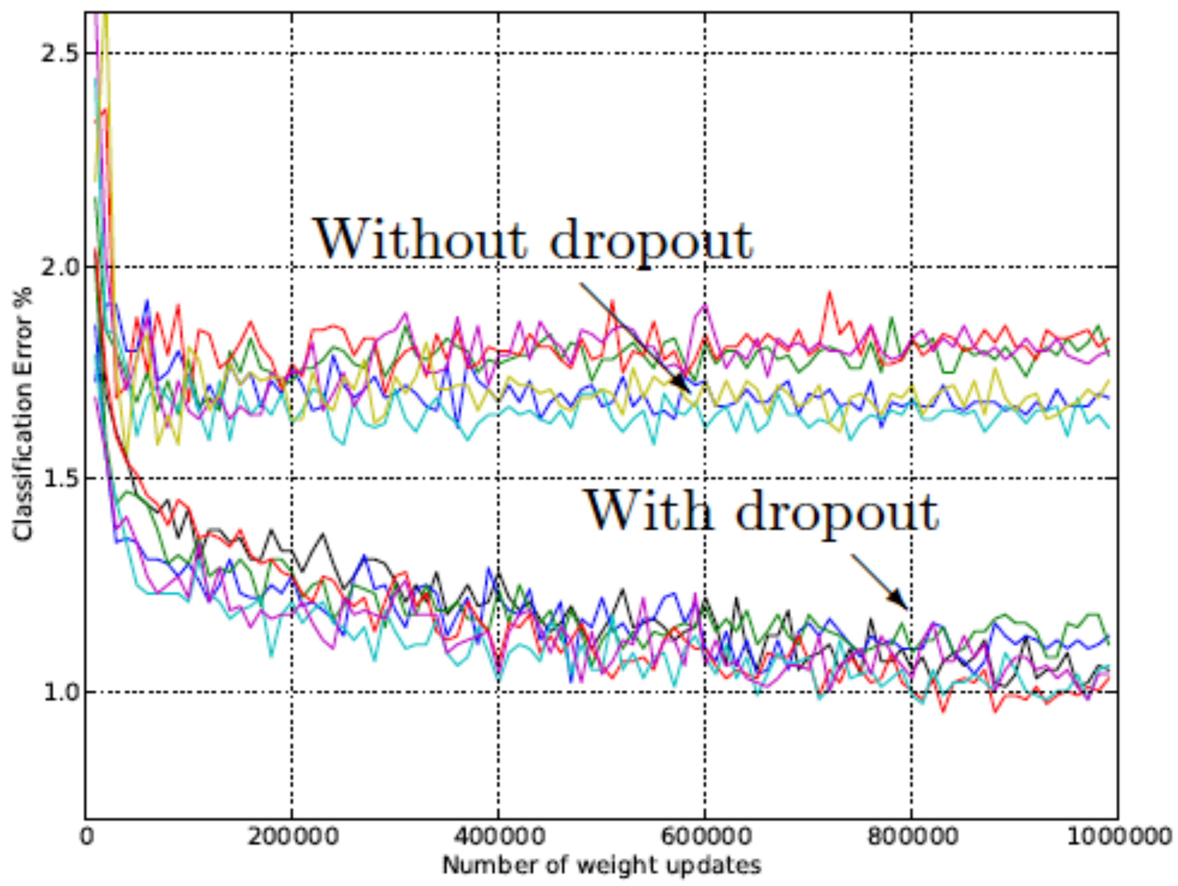


Figure 4: Test error for different architectures with and without dropout. The networks have 2 to 4 hidden layers each with 1024 to 2048 units.