Deep Learning: Theory and Practice

Deep Learning - Practical Considerations

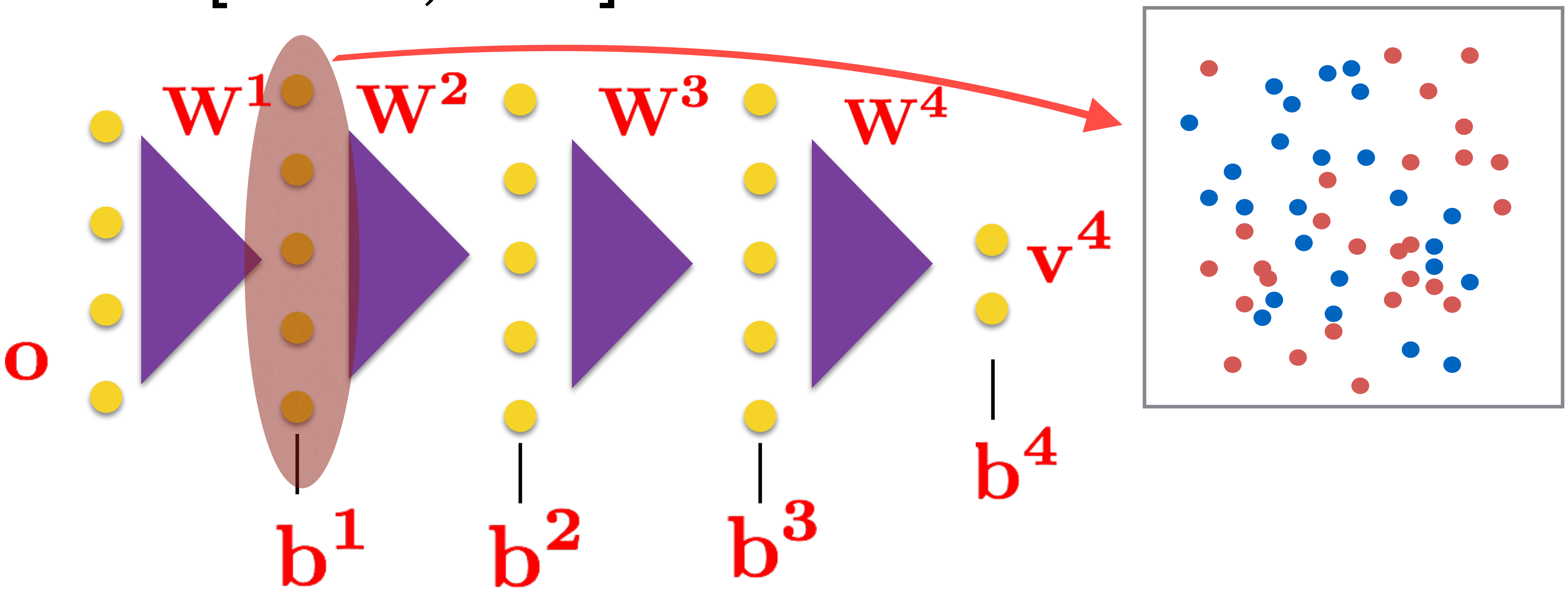
02-04-2020

deeplearning.cce2020@gmail.com

Established 1911

Deep Networks Intuition

Neural networks with multiple hidden layers - Deep networks [Hinton, 2006]

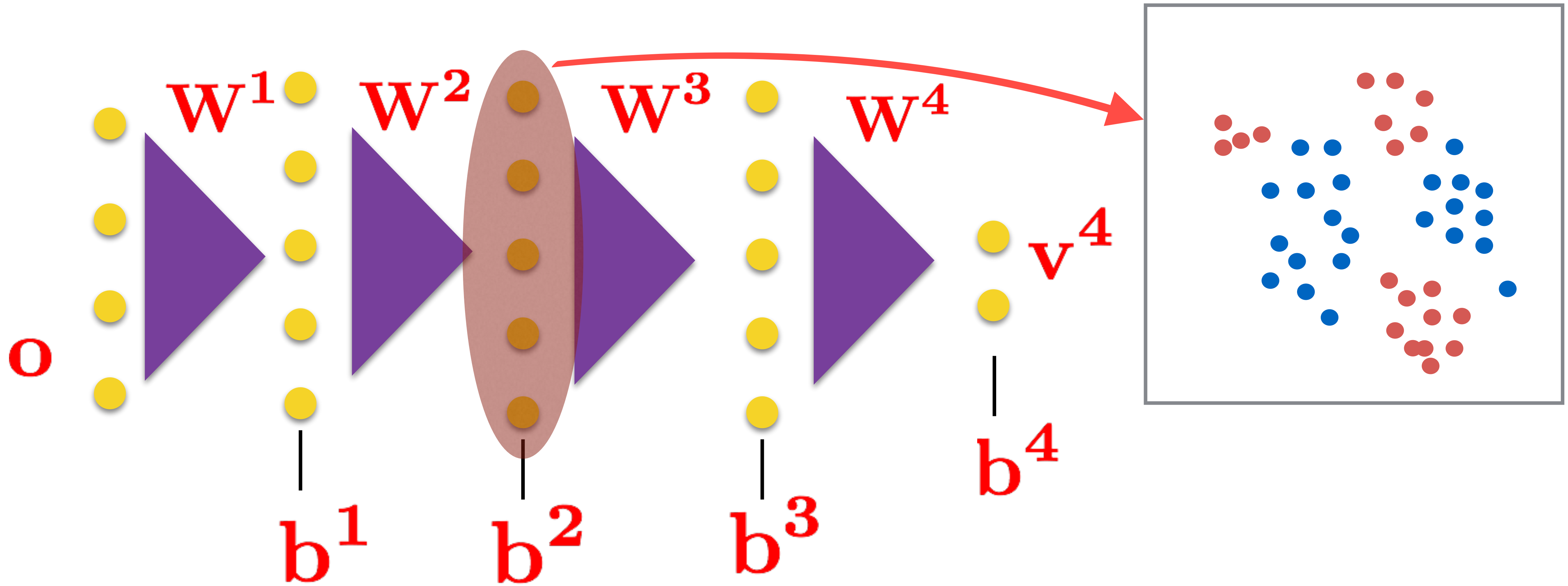


Established 1911

INDIAN INSTITUTE OF SCIENCE
भारतीय विज्ञान संस्थान

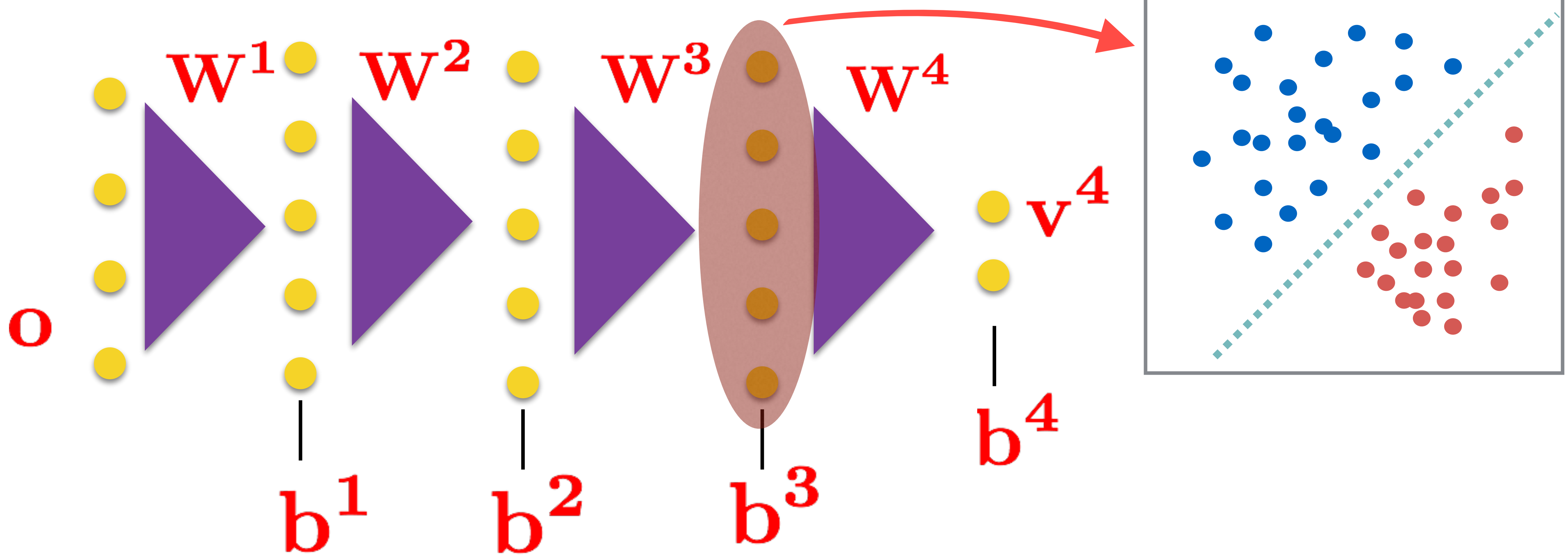
Deep Networks Intuition

Neural networks with multiple hidden layers - Deep networks



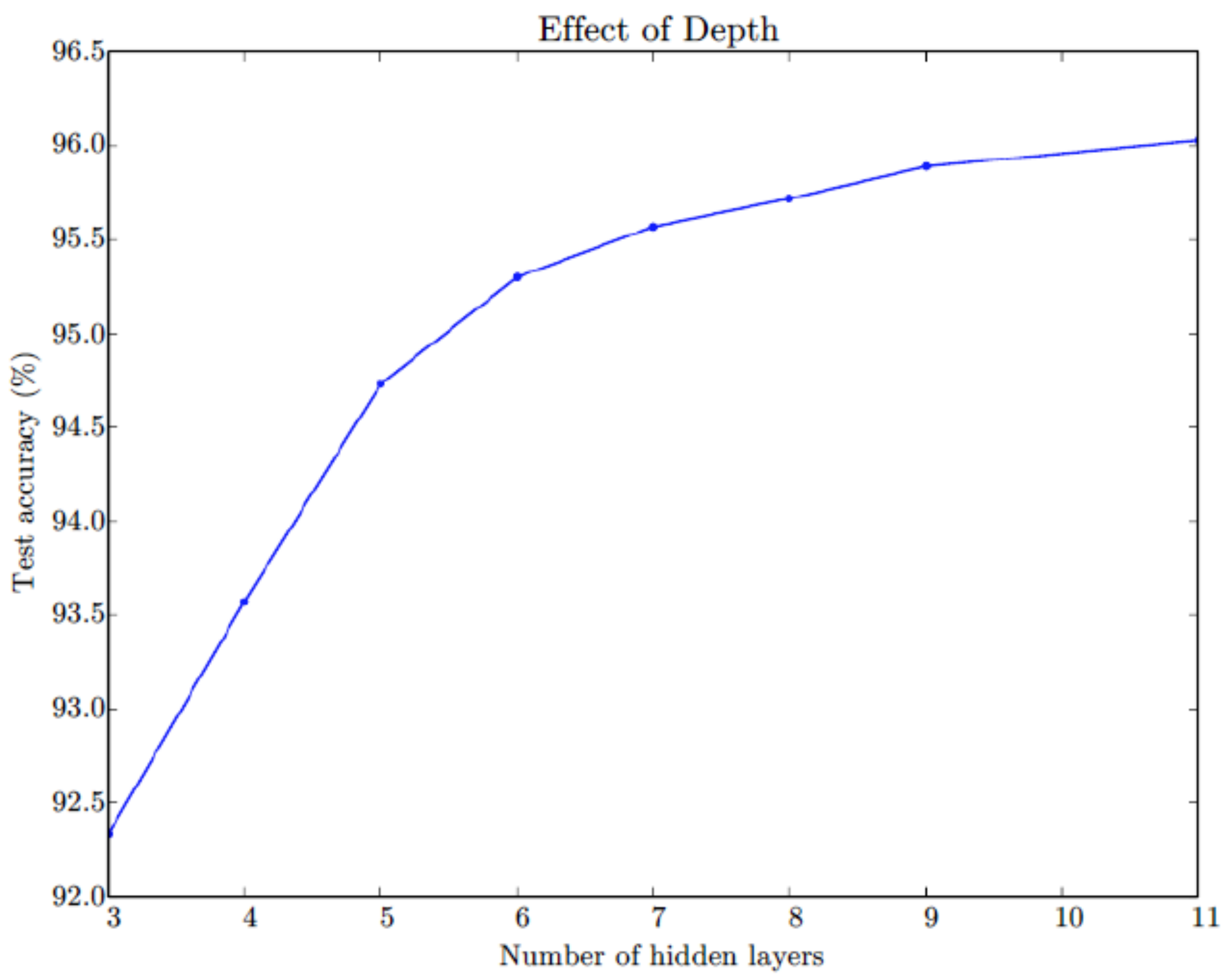
Deep Networks Intuition

Neural networks with multiple hidden layers - Deep networks



Deep networks perform hierarchical data abstractions which enable the non-linear separation of complex data samples.

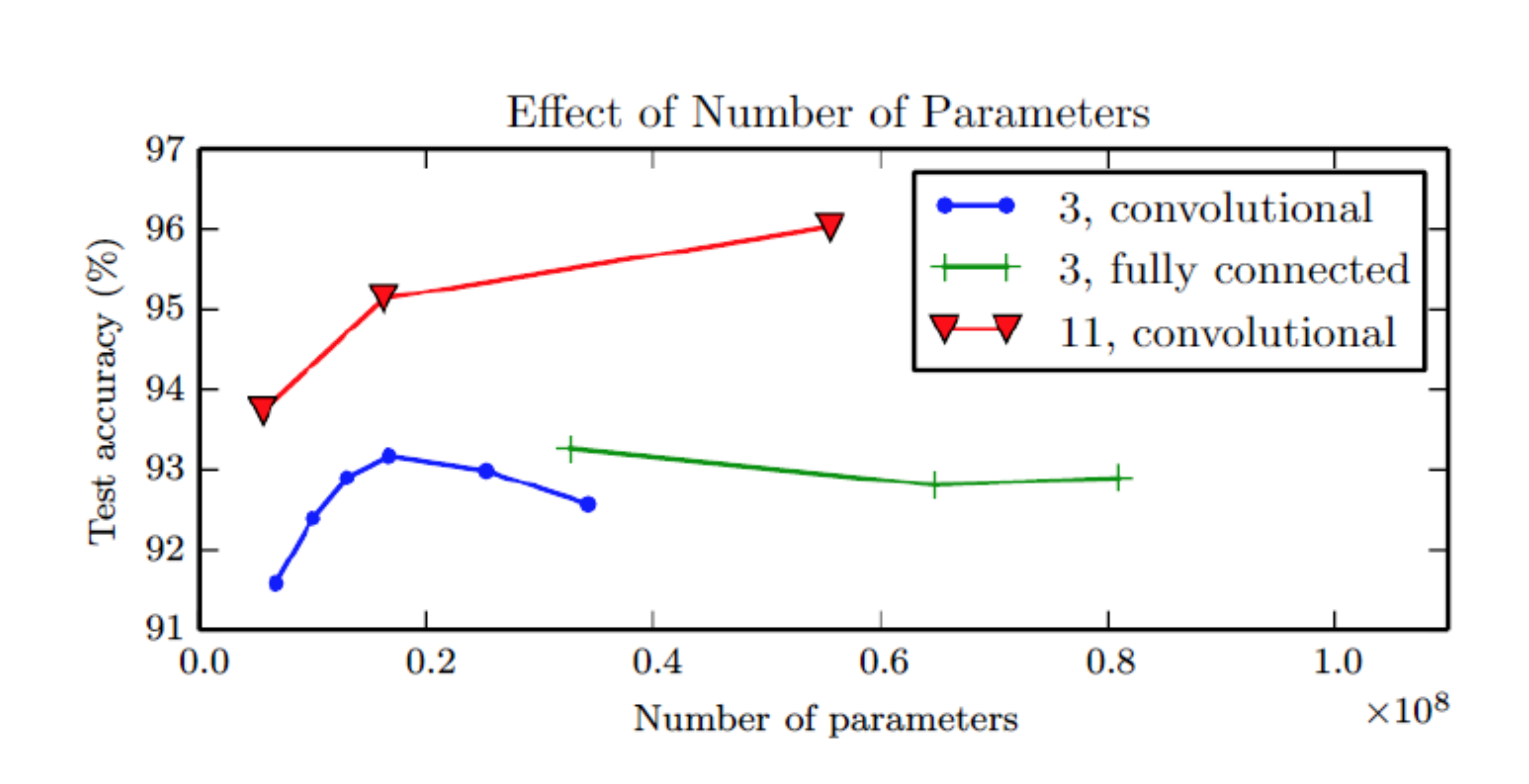
Need for Depth



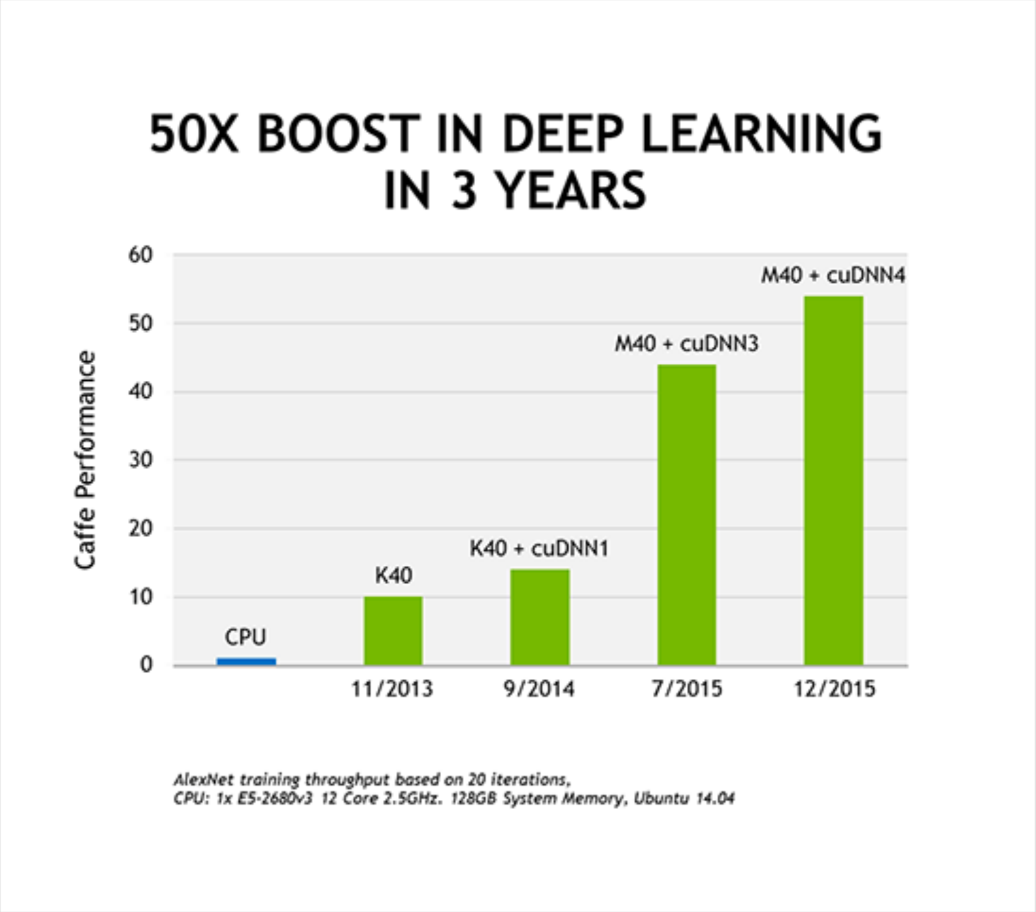
$$\boldsymbol{h}^{(1)} = g^{(1)}\left(\boldsymbol{W}^{(1)\top}\boldsymbol{x} + \boldsymbol{b}^{(1)}\right)$$

$$\boldsymbol{h}^{(2)} = g^{(2)}\left(\boldsymbol{W}^{(2)\top}\boldsymbol{h}^{(1)} + \boldsymbol{b}^{(2)}\right)$$

Need for Depth



Deep Networks



- Are these networks trainable ?
  - Advances in computation and processing
  - Graphical processing units (GPUs) performing multiple parallel multiply accumulate operations.
  - Large amounts of supervised data sets

Established 1911

Deep Networks

- Will the networks generalize with deep networks
  - DNNs are quite data hungry and performance improves by increasing the data.
  - Generalization problem is tackled by providing training data from all possible conditions.
    - Many artificial data augmentation methods have been successfully deployed
  - Providing the state-of-art performance in several real world applications.

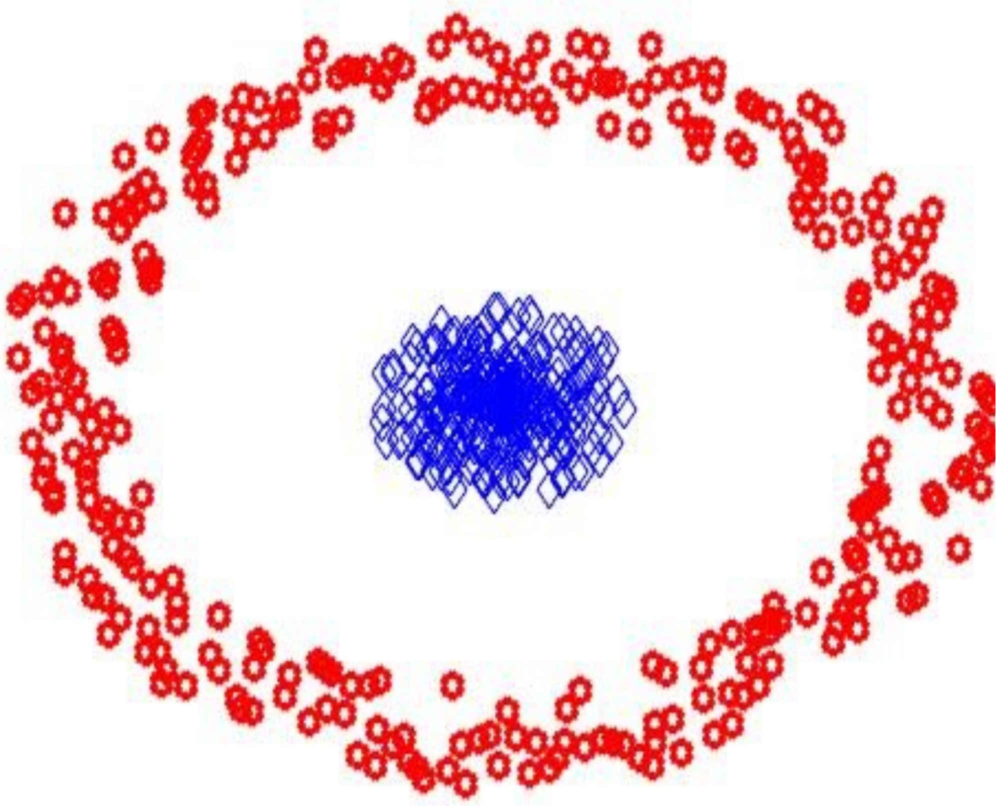
Established 1911

INDIAN INSTITUTE OF SCIENCE
भारतीय विज्ञान संस्थान

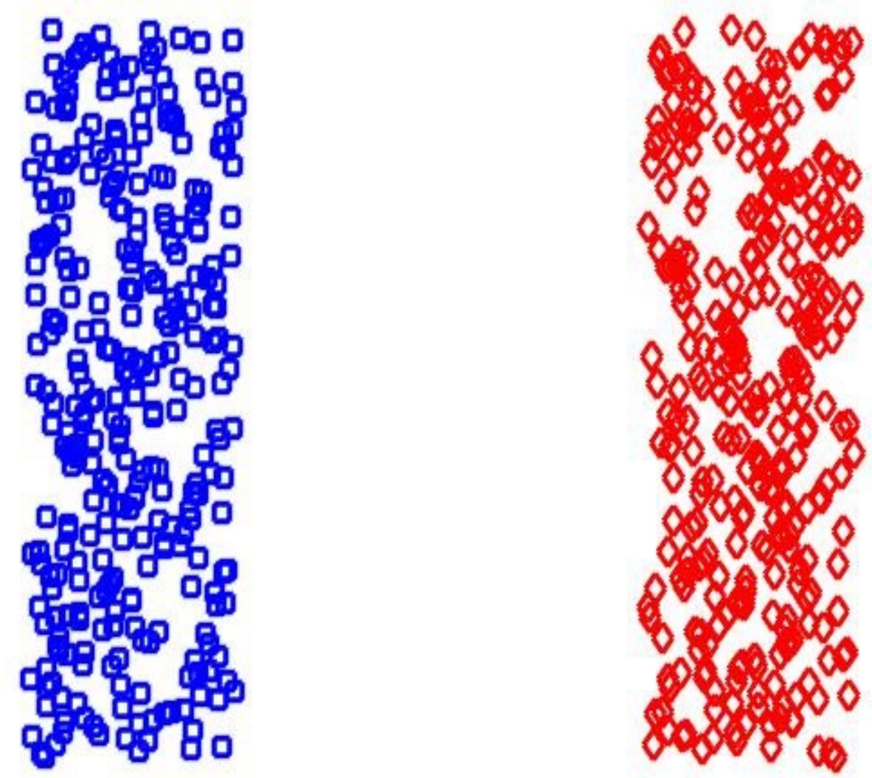
Representation Learning in Deep Networks

- The input data representation is one of most important components of any machine learning system.

Cartesian Coordinates



Polar Coordinates



Established 1911

Representation Learning in Deep Networks

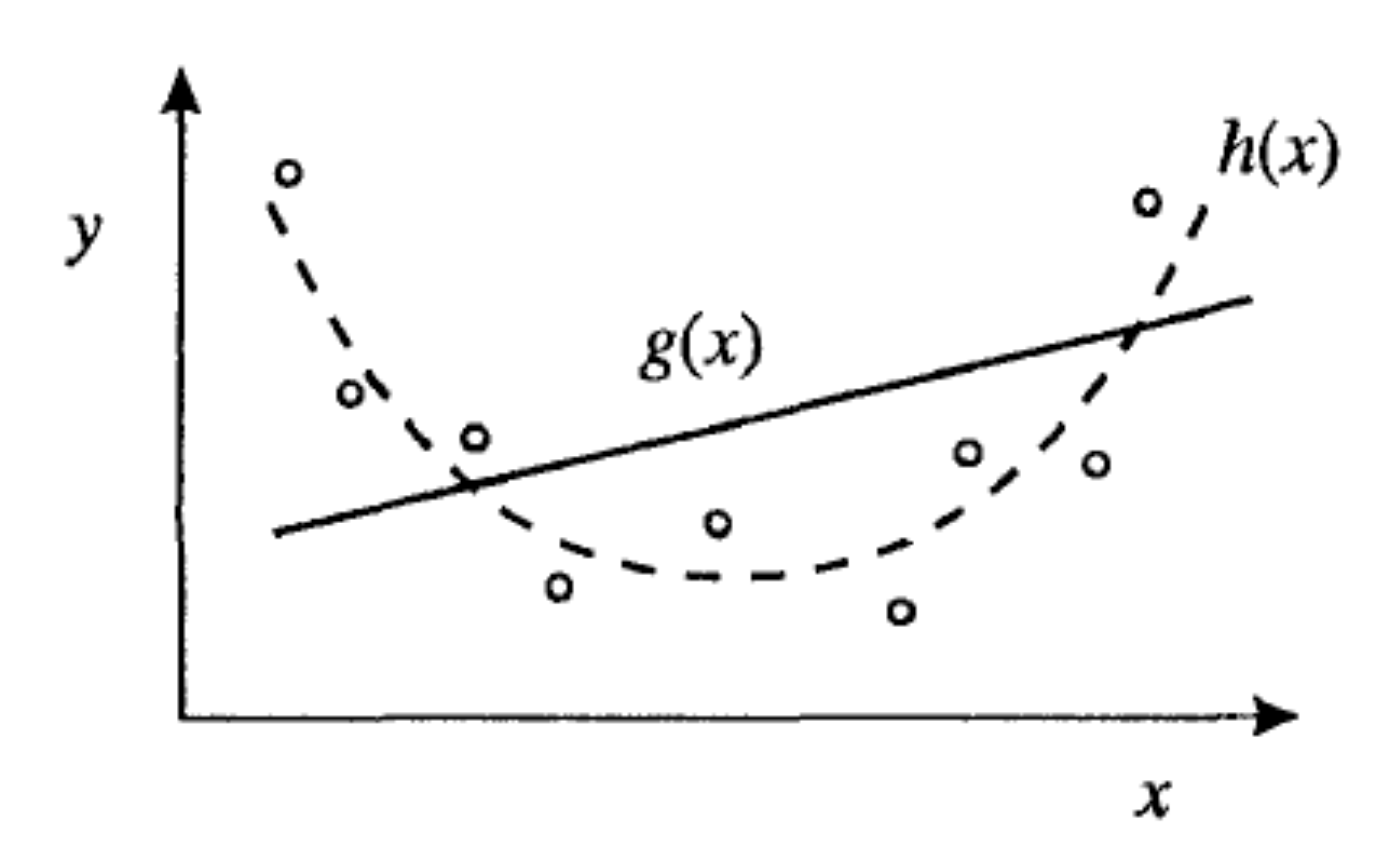
- The input data representation is one of most important components of any machine learning system.
  - Extract factors that enable classification while suppressing factors which are susceptible to noise.

- Finding the right representation for real world applications - substantially challenging.
  - Deep learning solution - build complex representations from simpler representations.
  - The dependencies between these hierarchical representations are refined by the target.

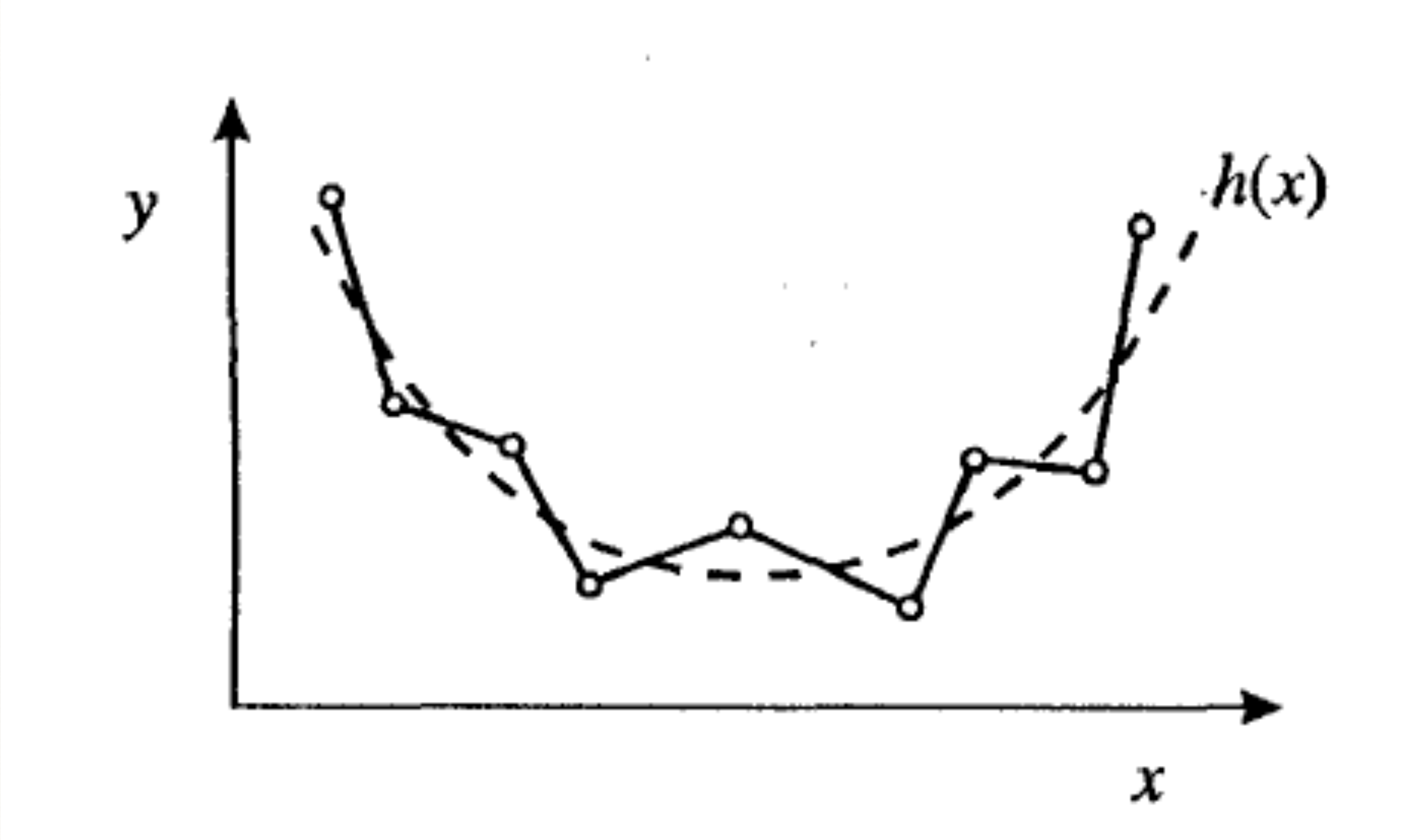
Established 1911

INDIAN INSTITUTE OF SCIENCE
भारतीय विज्ञान संस्थान

Underfit



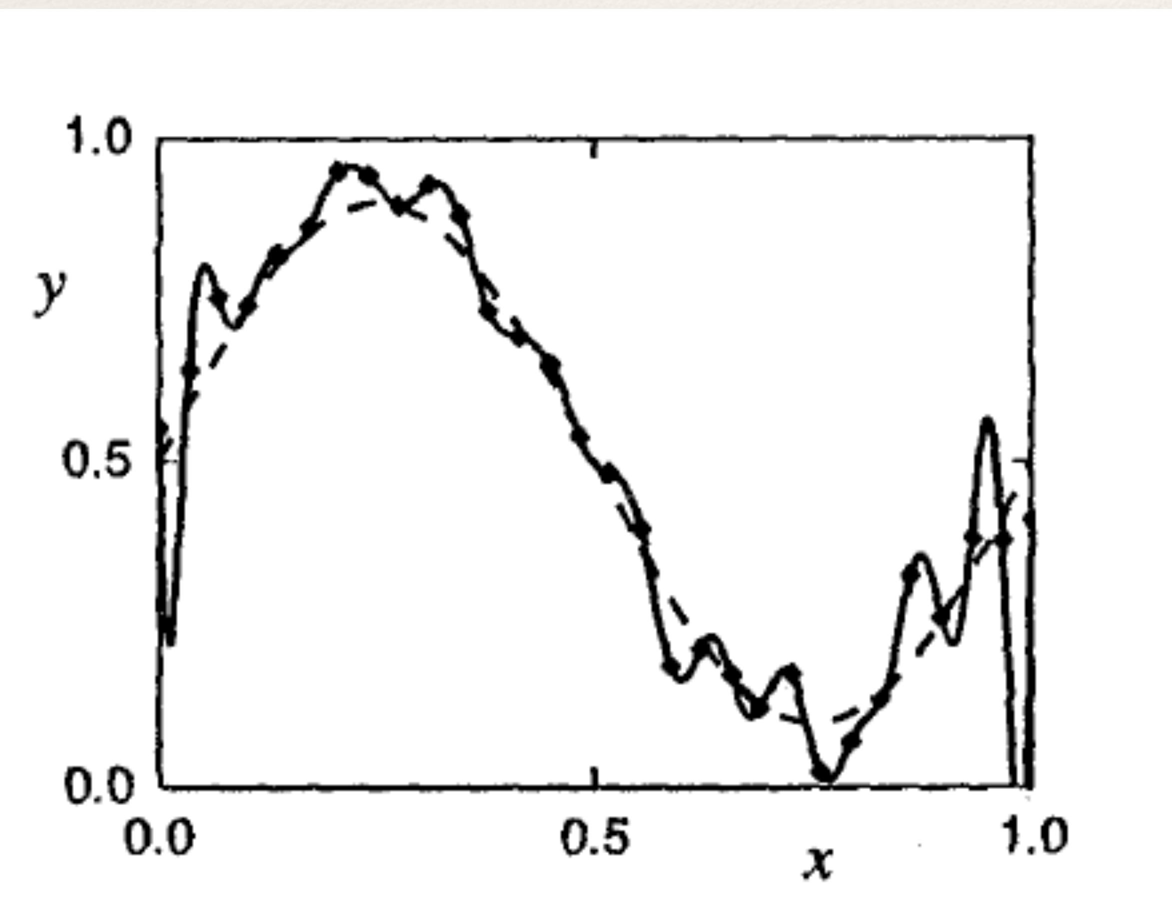
Overfit



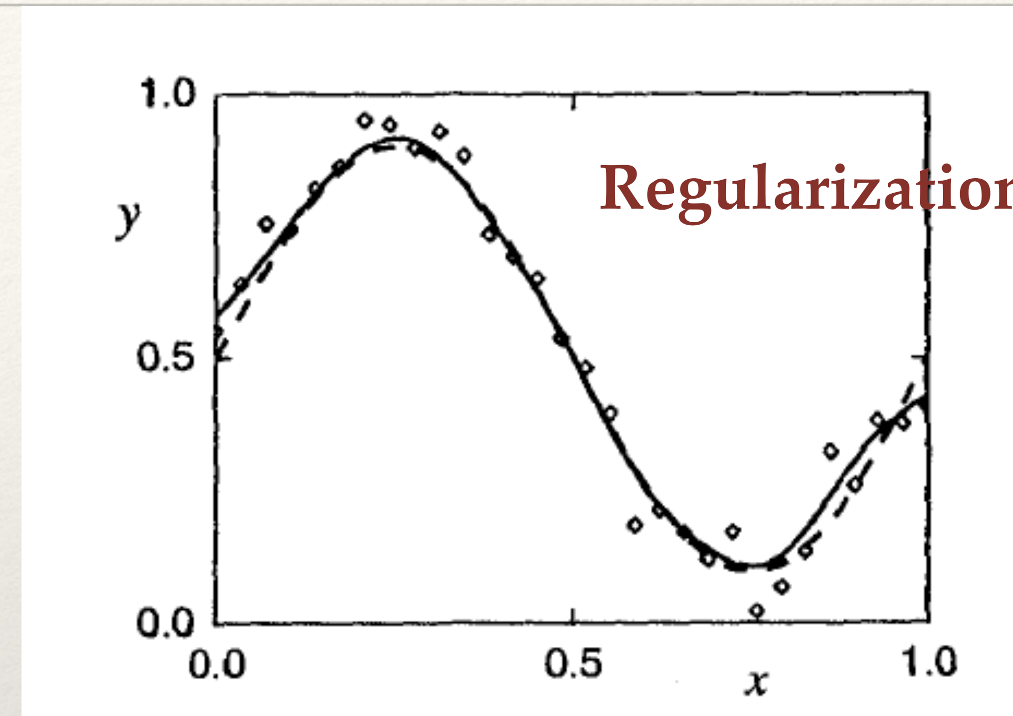
Avoiding OverFitting In Practice

Weight Decay Regularization

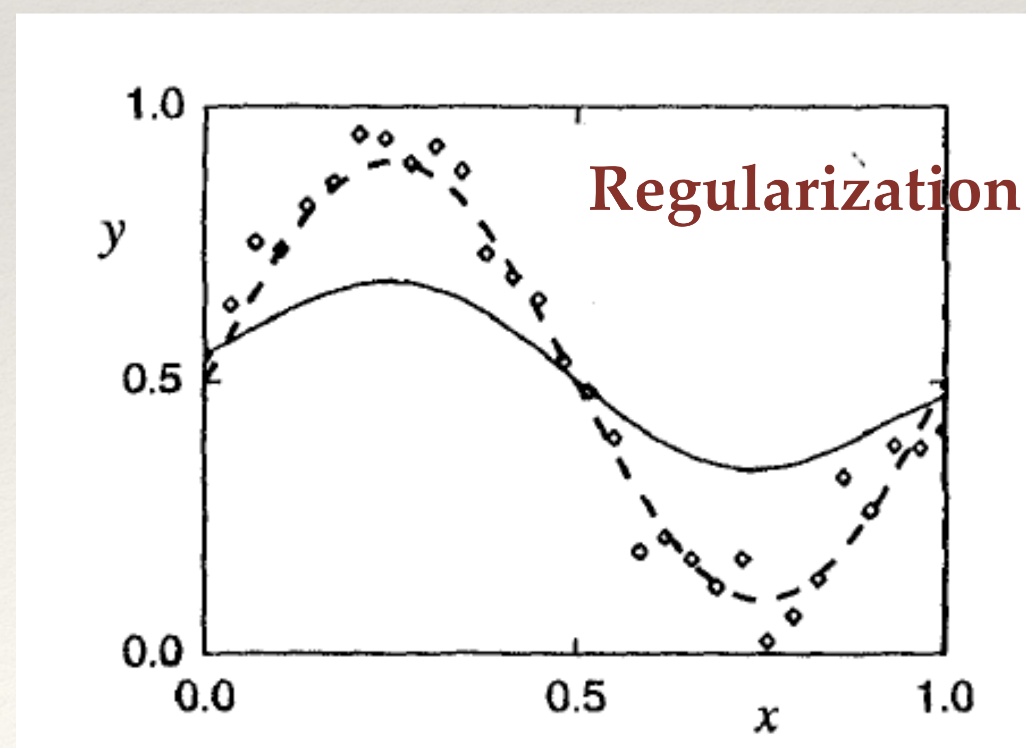
Regularization = 0



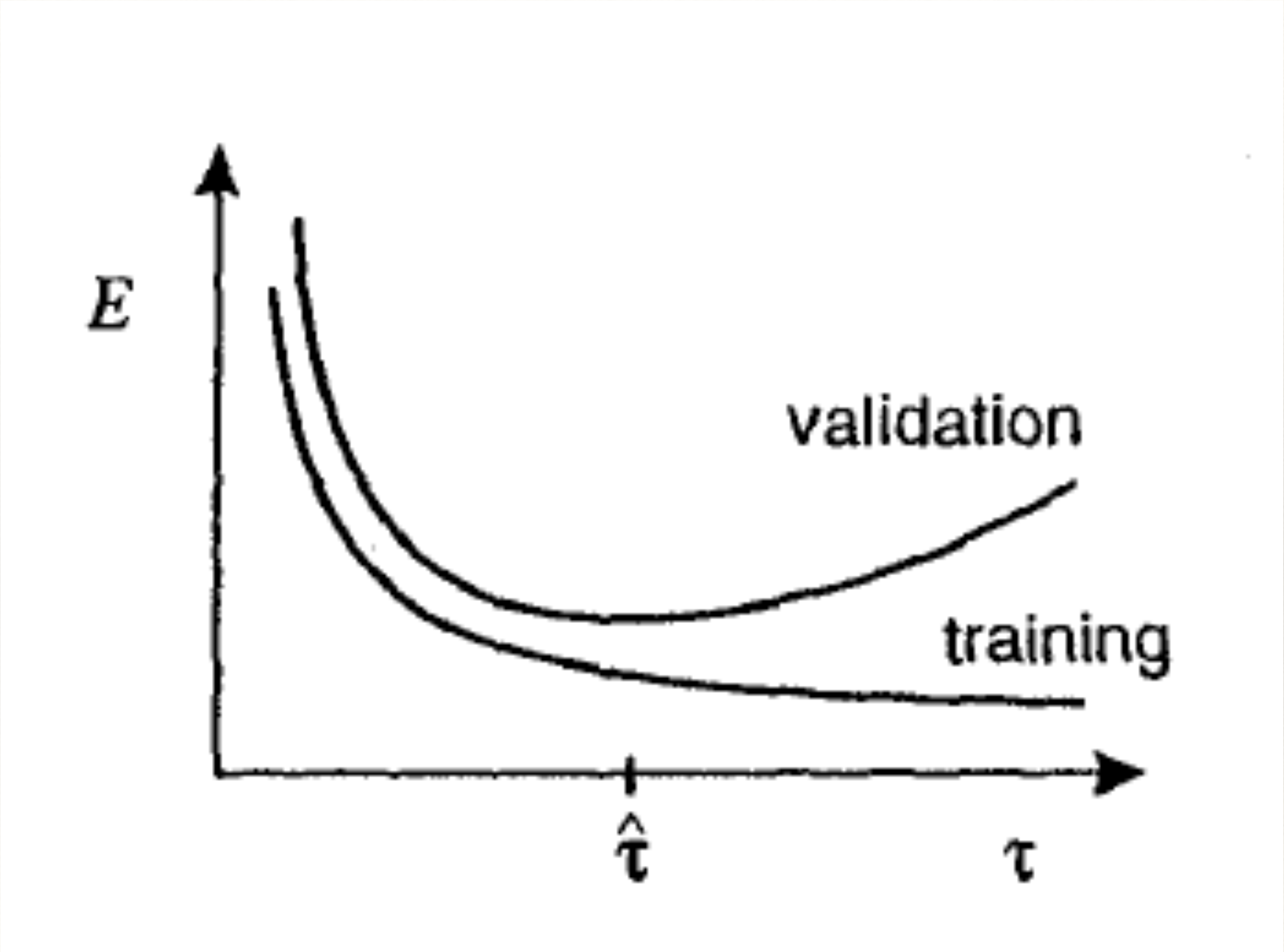
Regularization = 40



Regularization = 4000



Early Stopping



Most Popular in Practice

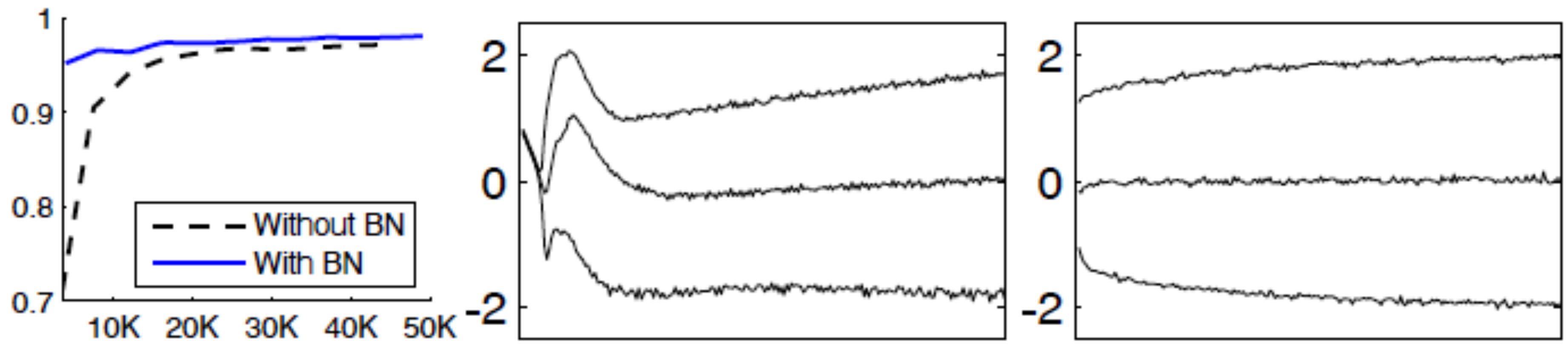
Batch Normalization

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.*

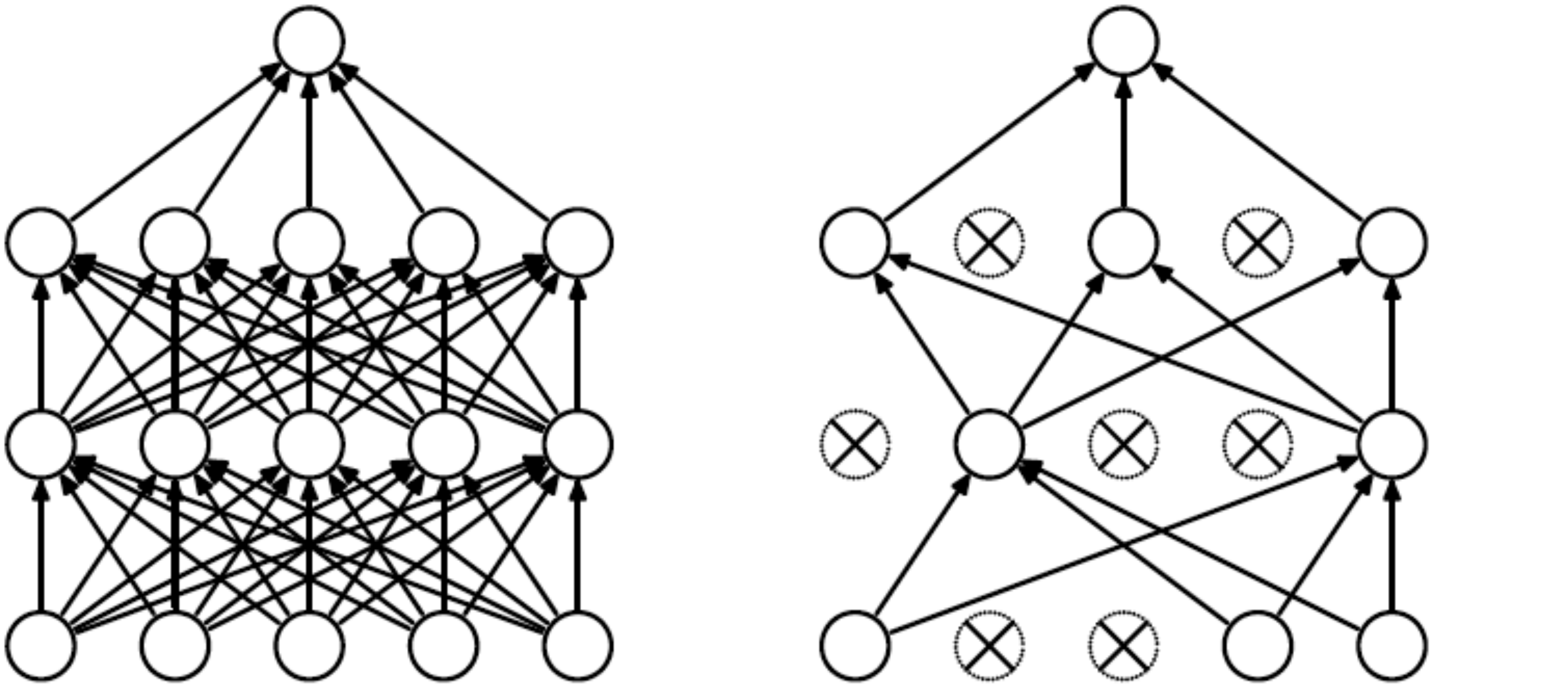
Dropout Strategy in Neural Network Training

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

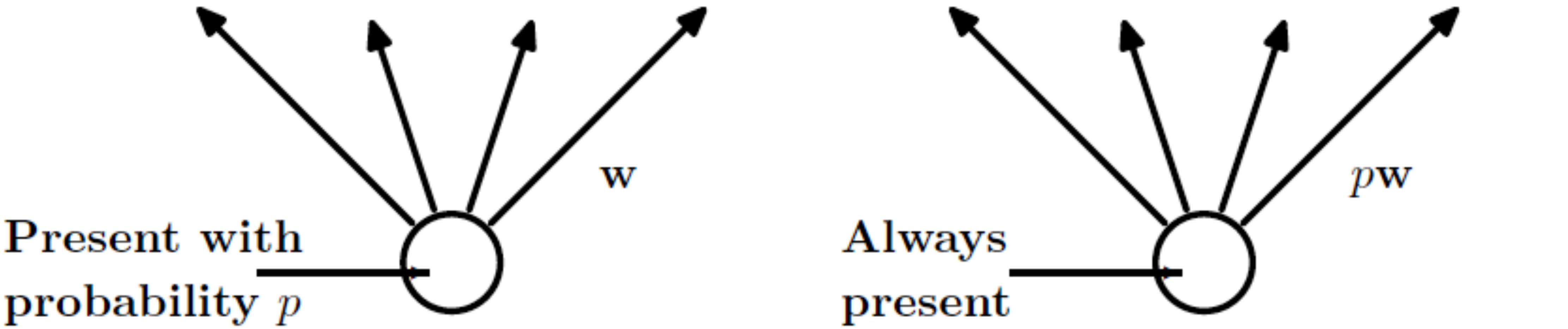
Dropouts in Neural Networks



(a) Standard Neural Net

(b) After applying dropout.

Dropout in Training and Test



(a) At training time

(b) At test time

Dropout Application

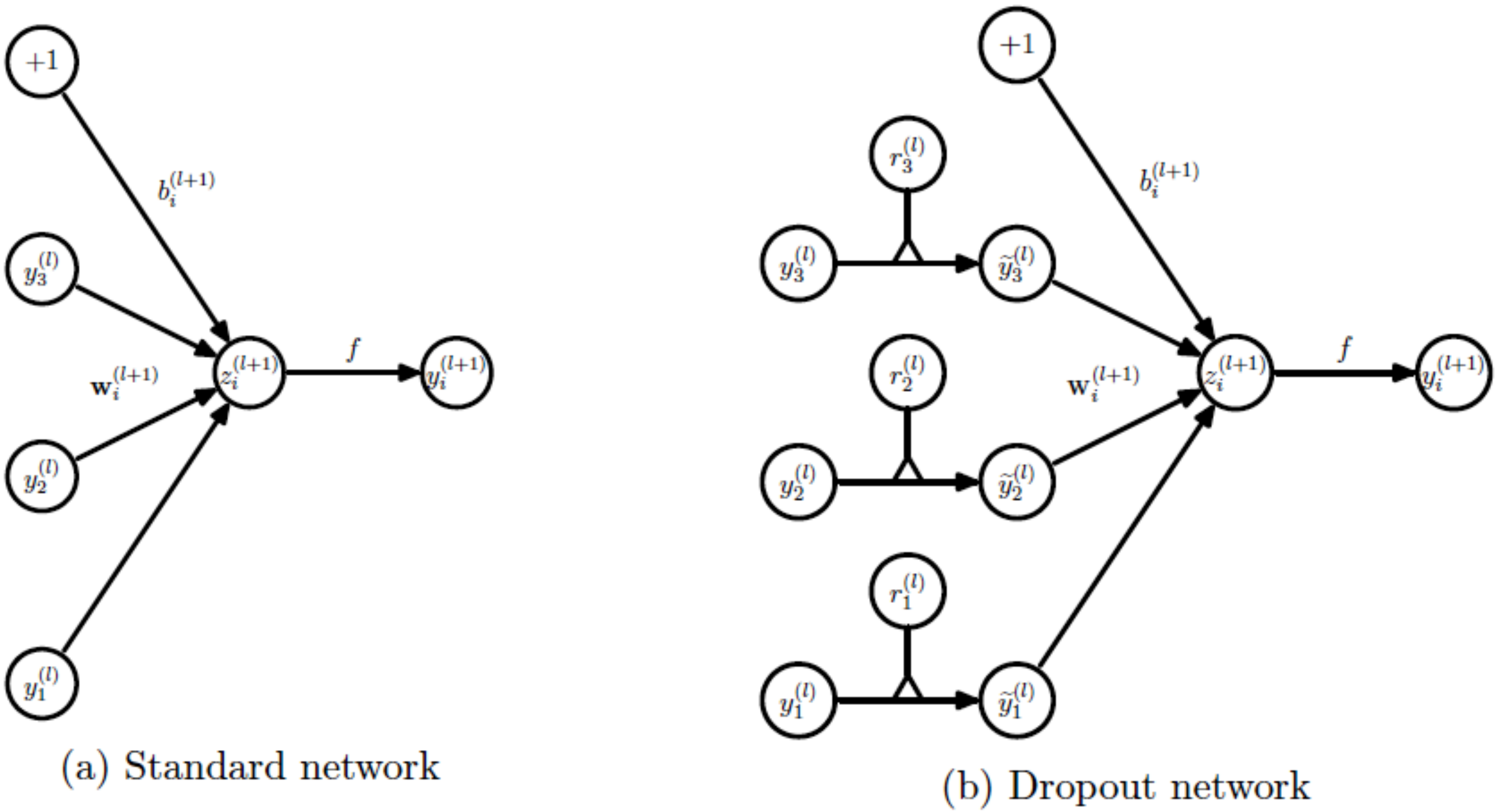


Figure 3: Comparison of the basic operations of a standard and dropout network.

Effect of Dropouts

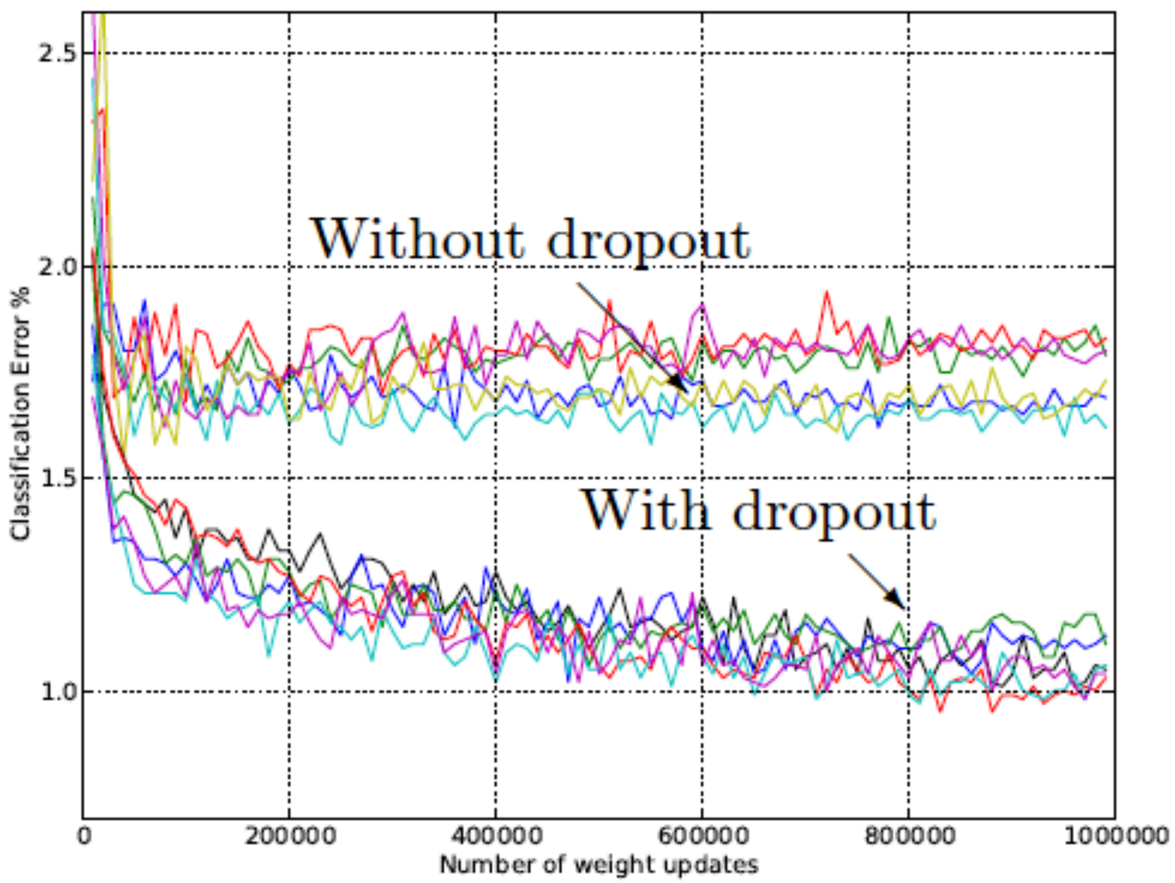
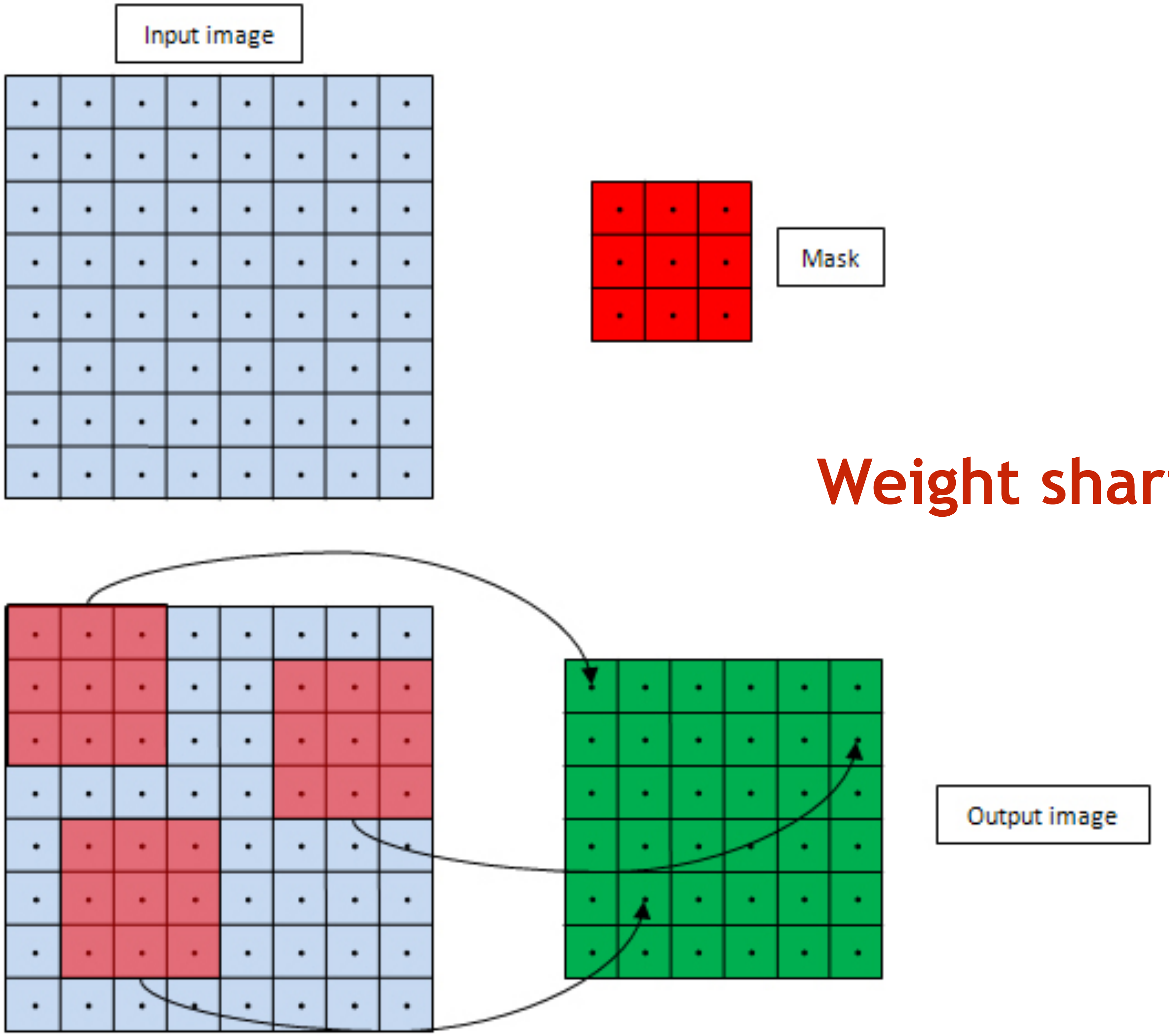


Figure 4: Test error for different architectures with and without dropout. The networks have 2 to 4 hidden layers each with 1024 to 2048 units.

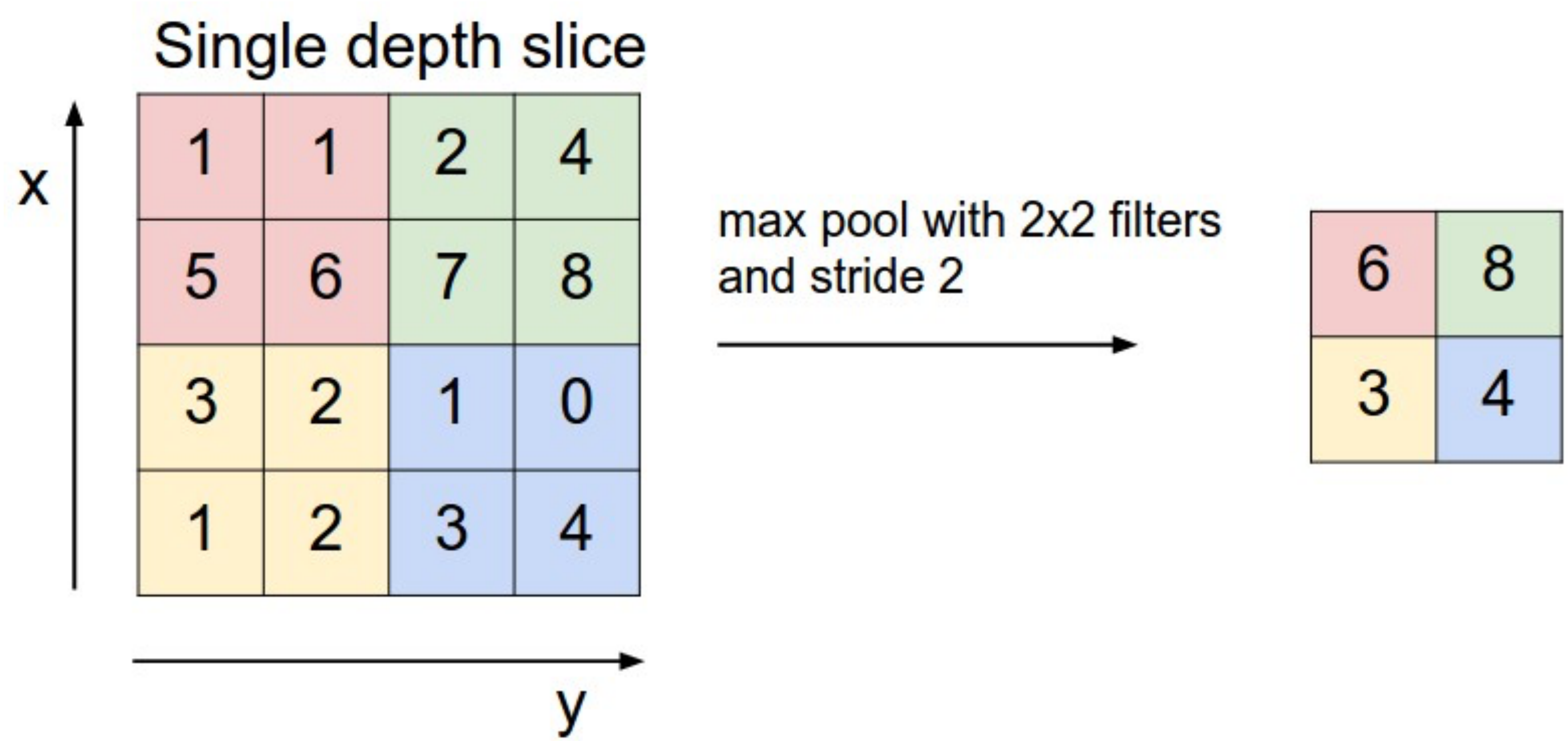
Convolutional Neural Networks

Other Architectures - Convolution Operation

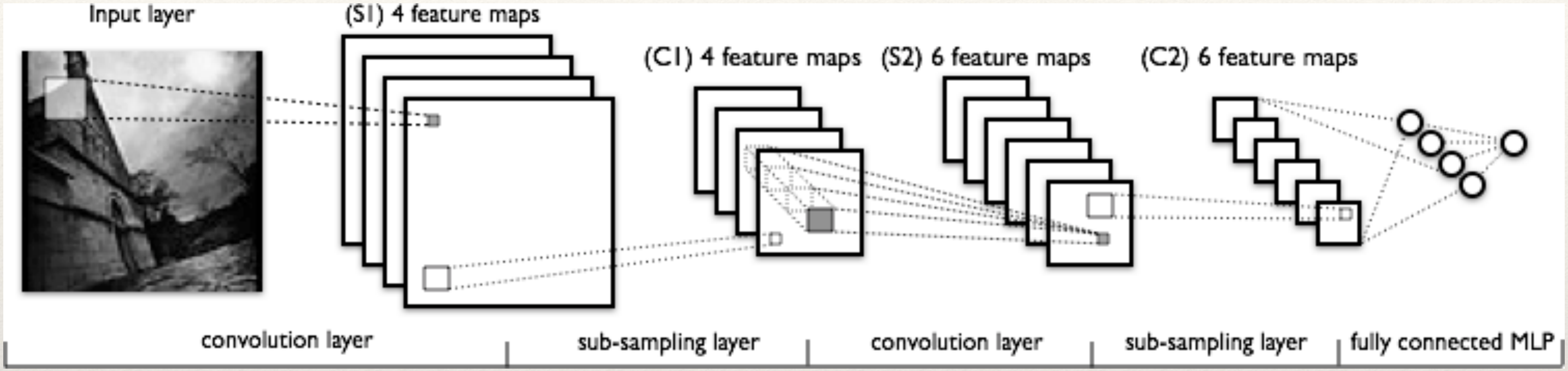


Weight sharing

Max Pooling Operation

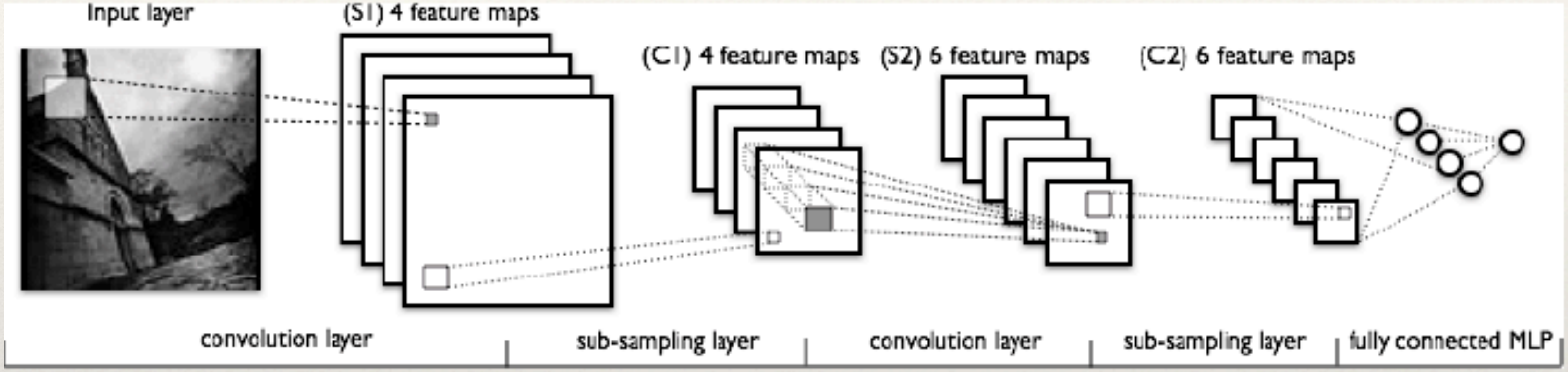


Convolutional Neural Networks



- Multiple levels of filtering and subsampling operations.
- Feature maps are generated at every layer.

Convolutional Neural Networks



- Multiple levels of filtering and subsampling operations.
- Feature maps are generated at every layer.

Back Propagation in CNNs