E9 205 Machine Learning for Signal Processing

Deep Neural Networks

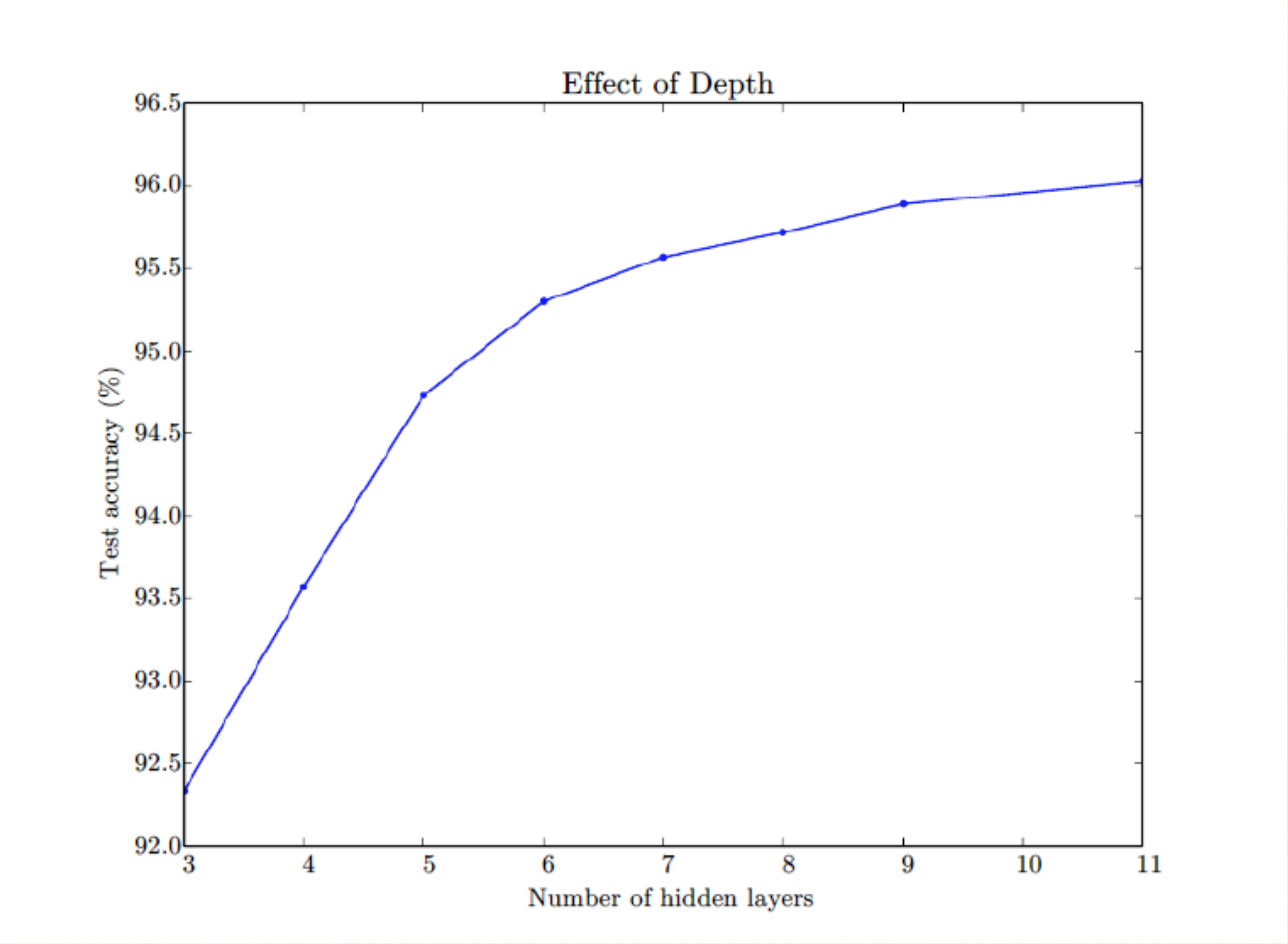
30-10-2019

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

Summary so far…

- Neural networks as discriminative classifiers
- Need for hidden layer
- Choice of non-linearities and target functions
- Estimating posterior probabilities with NNs
- Parameter learning with back propagation.

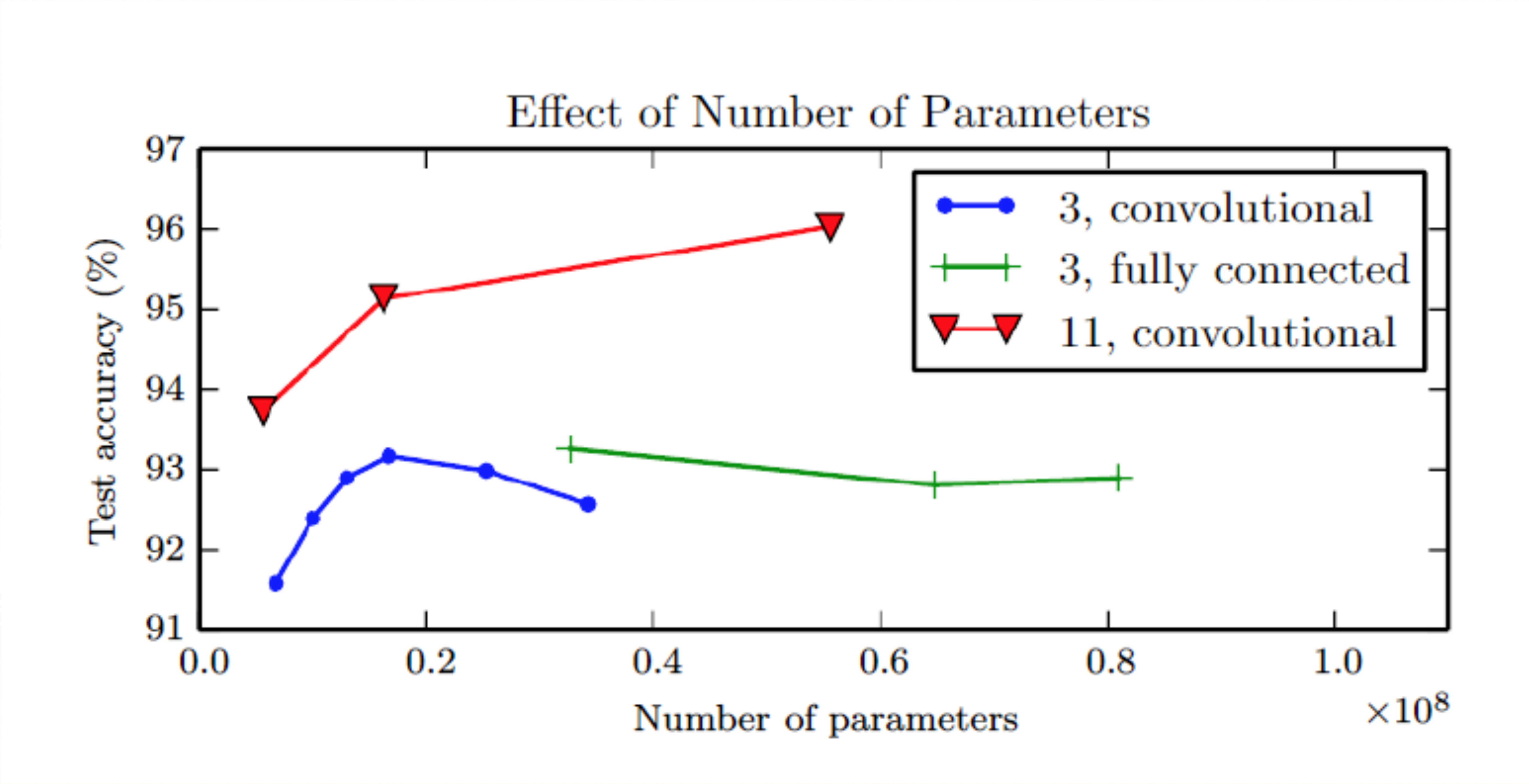
Need for Depth



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

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

Need for Depth



Need For Deep Networks

Modeling complex real world data like speech, image, text

- Single hidden layer networks are too restrictive.
- Needs large number of units in the hidden layer and trained with large amounts of data.
- Not generalizable enough.

Networks with multiple hidden layers - deep networks (Open questions till 2005)

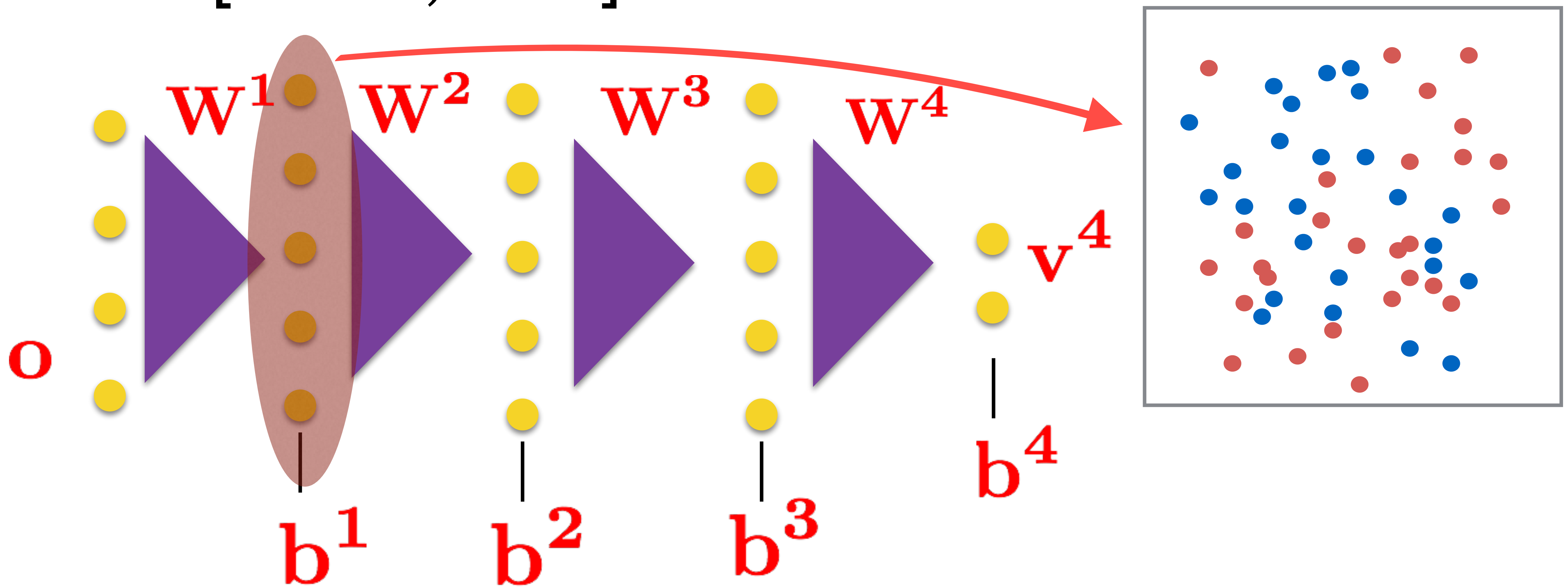
- Are these networks trainable ?
- How can we initialize such networks ?
- Will these generalize well or over train ?

Established 1911

INDIAN INSTITUTE OF SCIENCE
भारतीय विज्ञान संस्थान

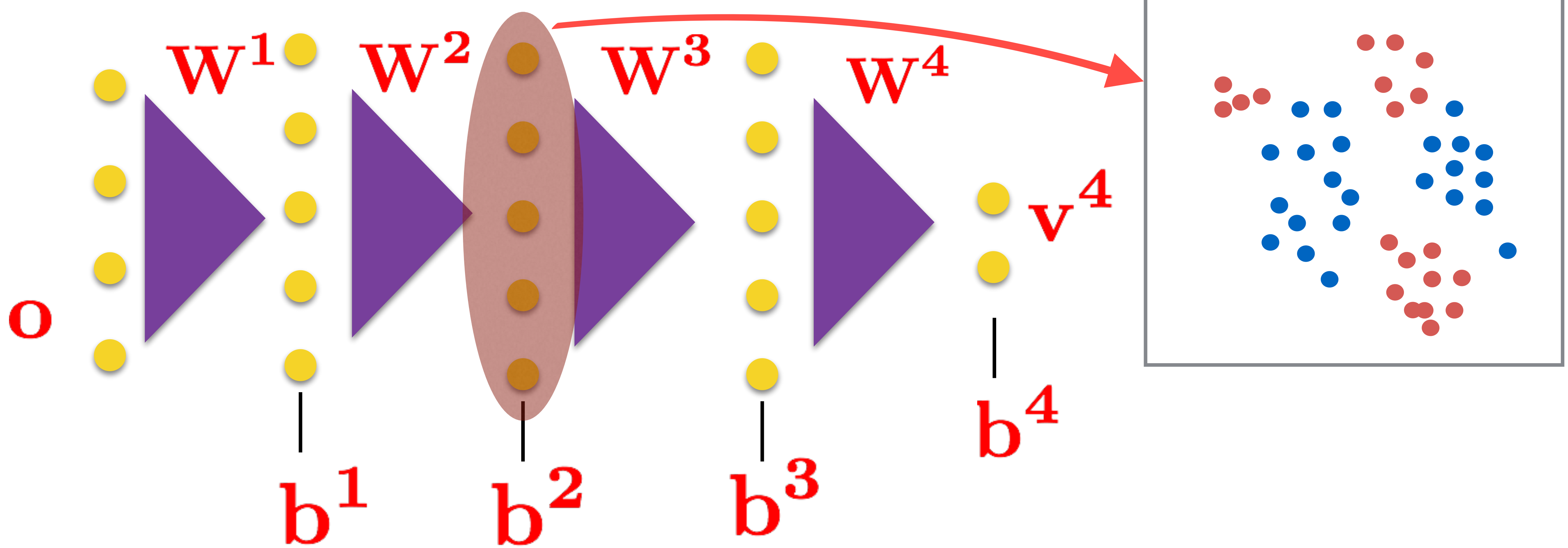
Deep Networks Intuition

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



Deep Networks Intuition

Neural networks with multiple hidden layers - Deep networks

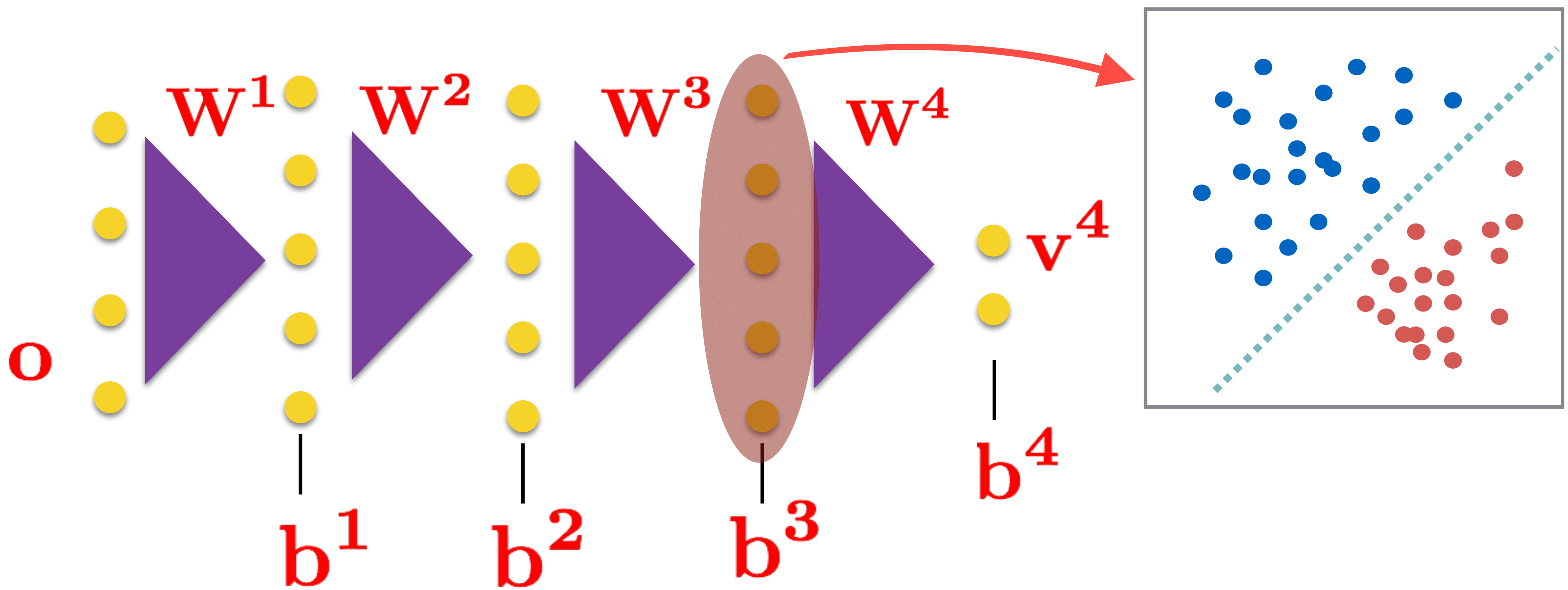


Established 1911

INDIAN INSTITUTE OF SCIENCE
भारतीय विज्ञान संस्थान

Deep Networks Intuition

Neural networks with multiple hidden layers - Deep networks

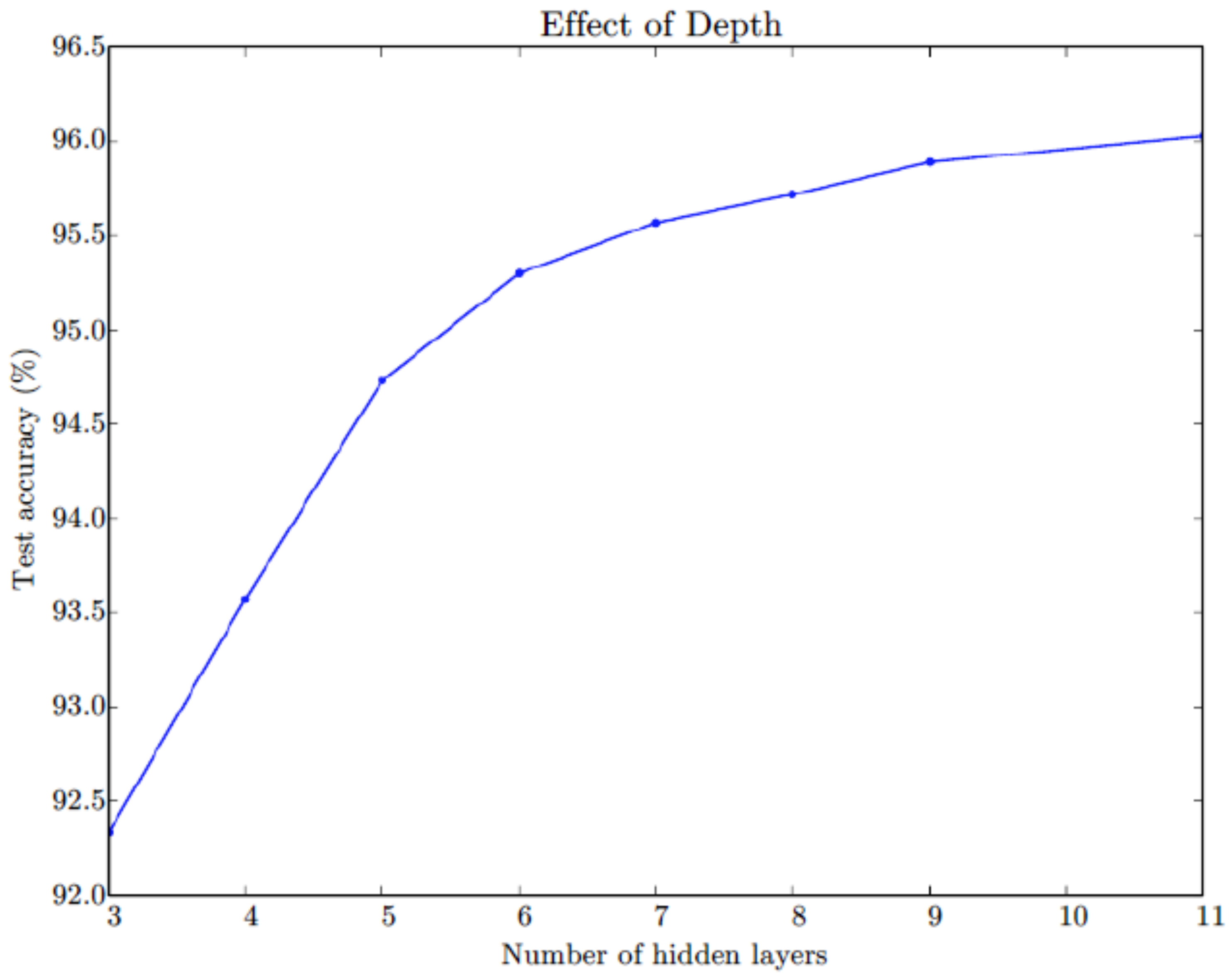


Deep networks perform hierarchical data abstractions which enable the non-linear separation of complex data samples.

Established 1911

INDIAN INSTITUTE OF SCIENCE
भारतीय विज्ञान संस्थान

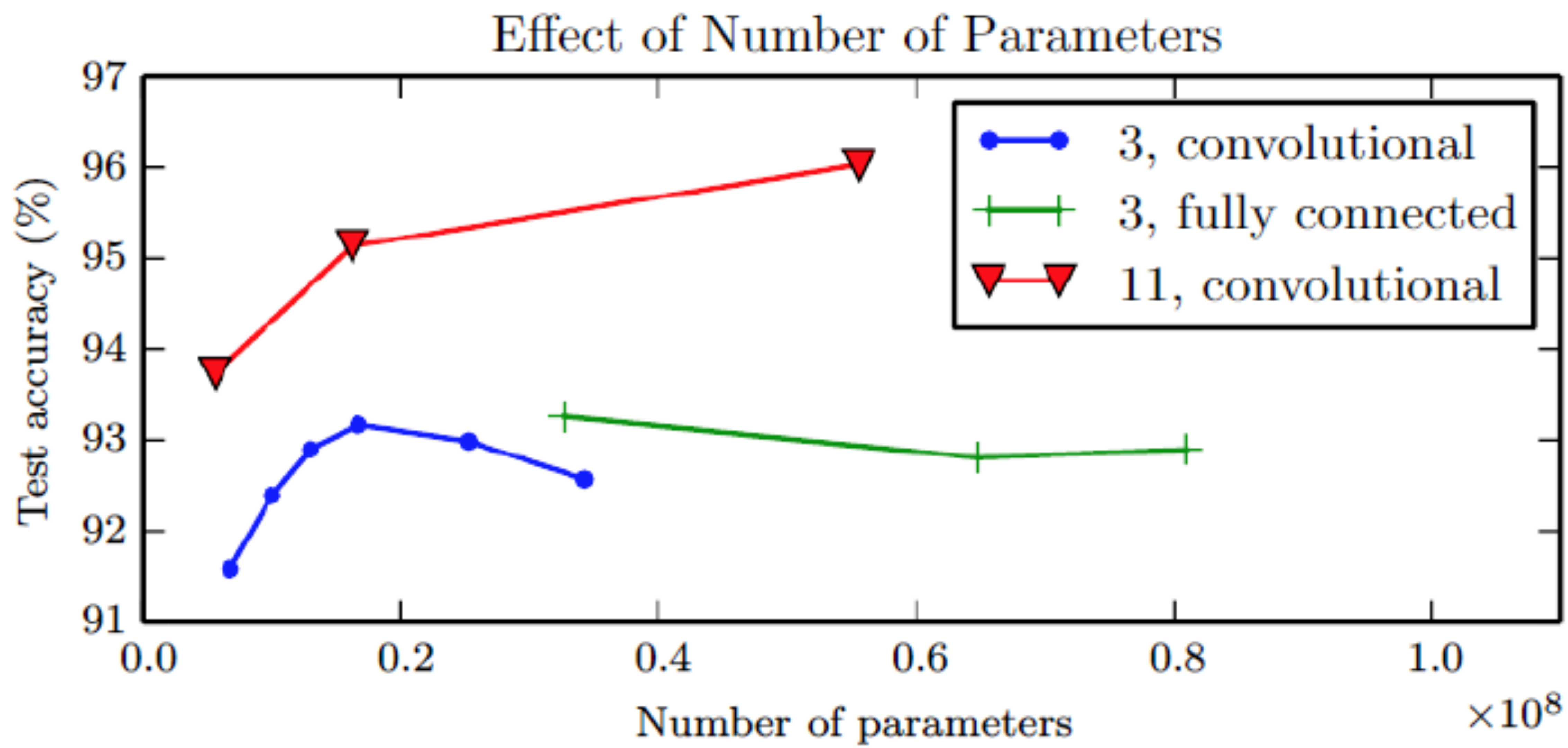
Need for Depth



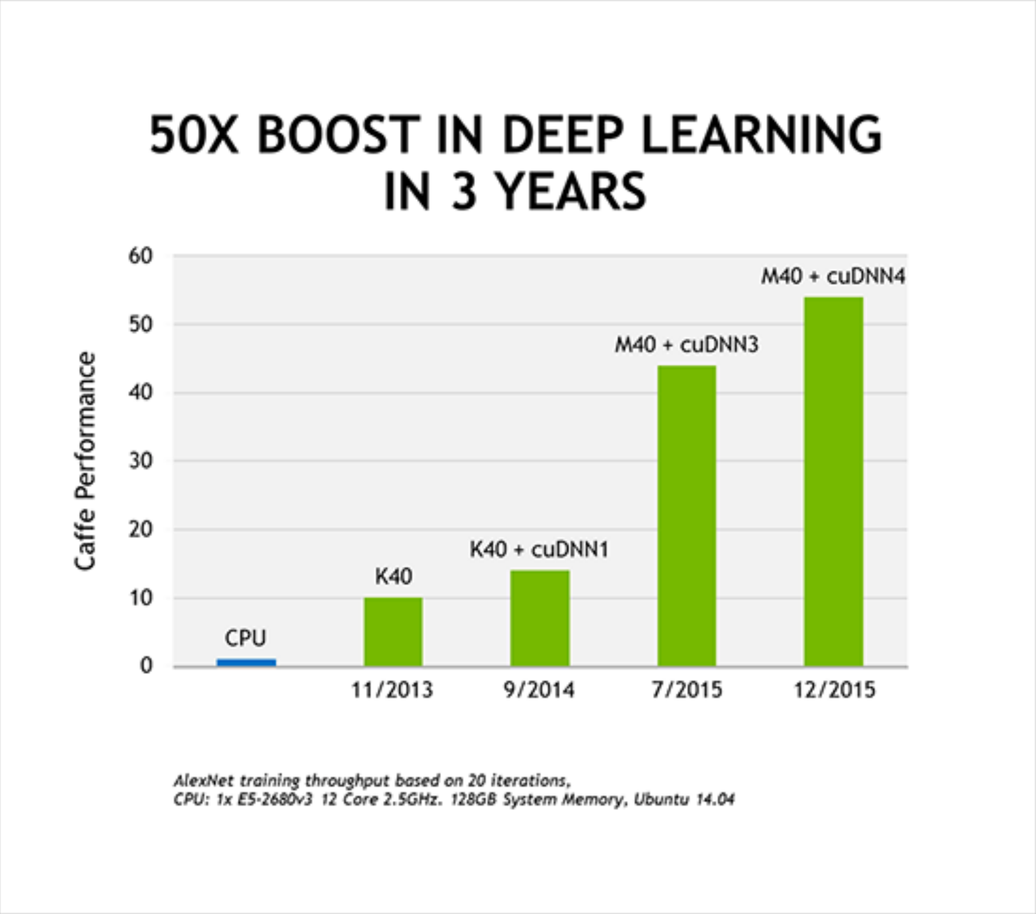
$$\boldsymbol{h}^{(1)} = g^{(1)}\left(\boldsymbol{W}^{(1)\top}\boldsymbol{x} + \boldsymbol{b}^{(1)}\right)$$

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

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

INDIAN INSTITUTE OF SCIENCE
भारतीय विज्ञान संस्थान

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.

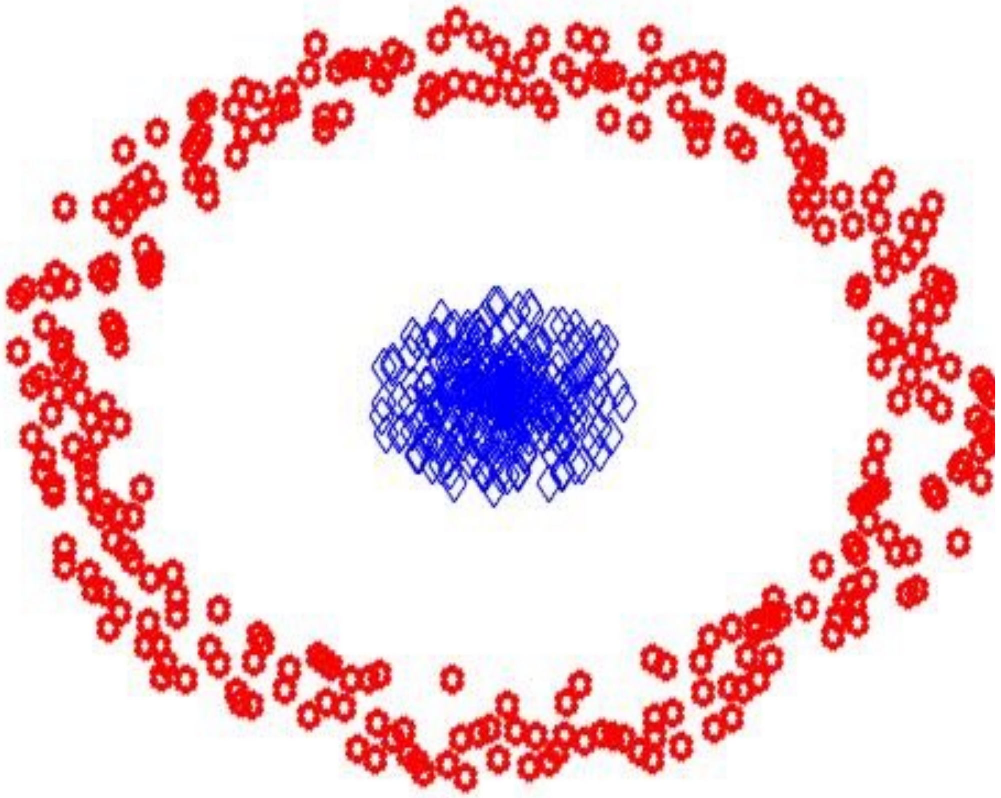
Established 1911

INDIAN INSTITUTE OF SCIENCE
भारतीय विज्ञान संस्थान

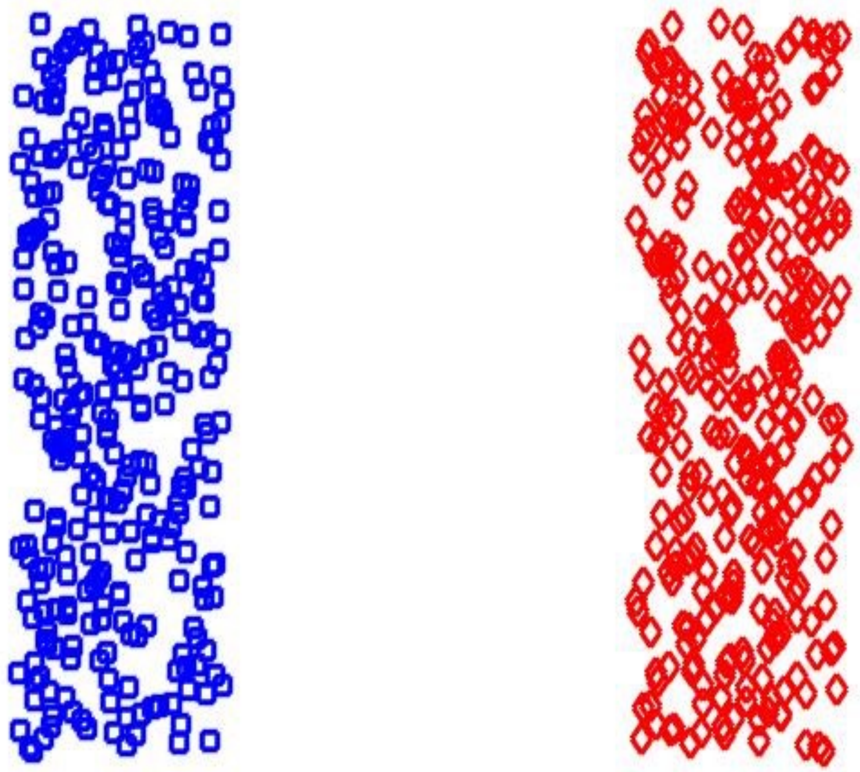
Representation Learning in Deep Networks

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

Cartesian Coordinates



Polar Coordinates



Representation Learning in Deep Networks

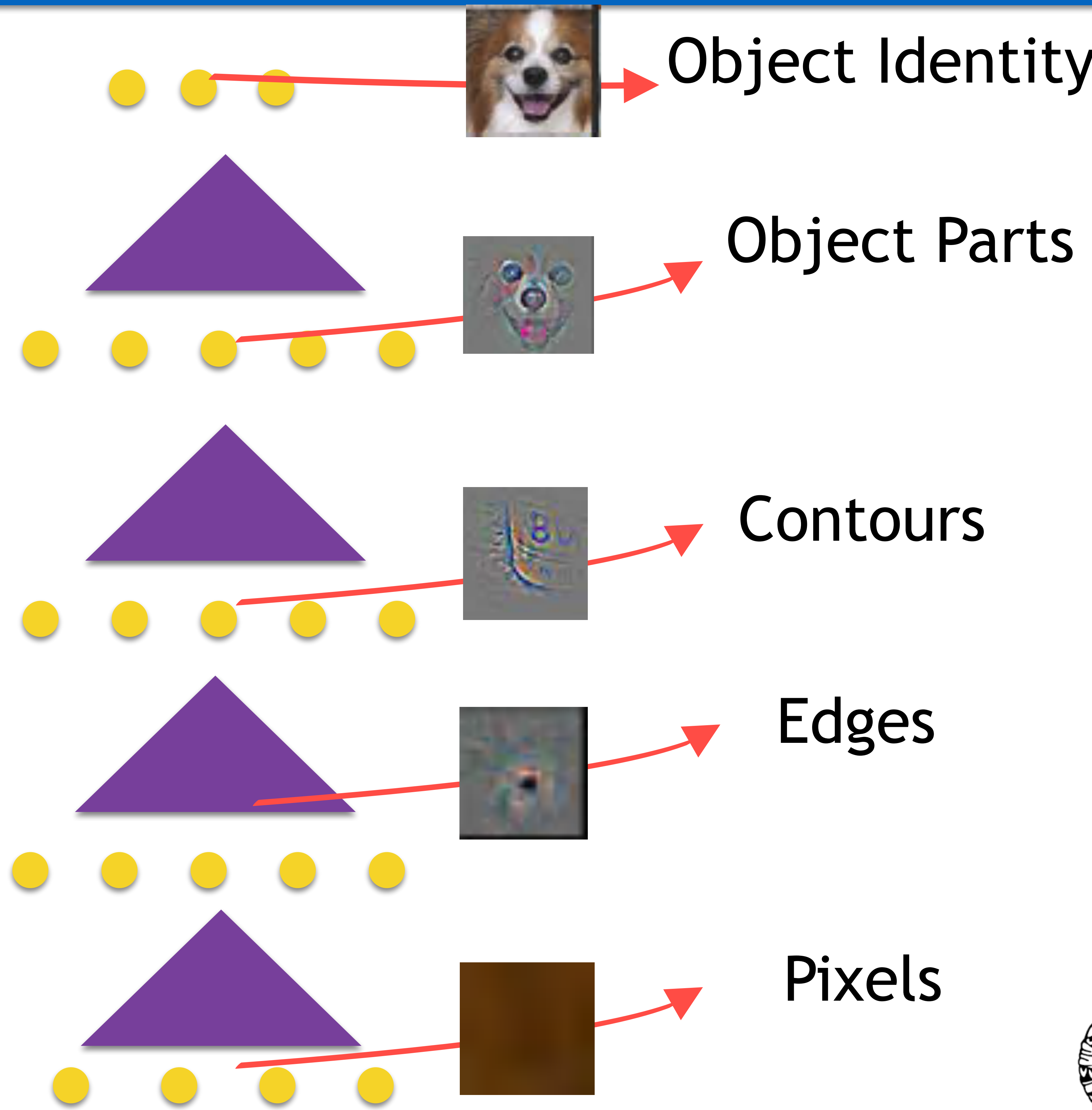
- The input data representation is one of most important components of any machine learning system.
  - Extract factors that enable classification while suppressing factors which are susceptible to noise.

- 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

Representation Learning in Deep Networks



[Zeiler, 2014]

On the Number of Linear Regions of Deep Neural Networks

Guido Montúfar
Max Planck Institute for Mathematics in the Sciences
montufar@mis.mpg.de

Razvan Pascanu
Université de Montréal
pascanur@iro.umontreal.ca

Kyunghyun Cho
Université de Montréal
kyunghyun.cho@umontreal.ca

Yoshua Bengio
Université de Montréal, CIFAR Fellow
yoshua.bengio@umontreal.ca

Deep Learning

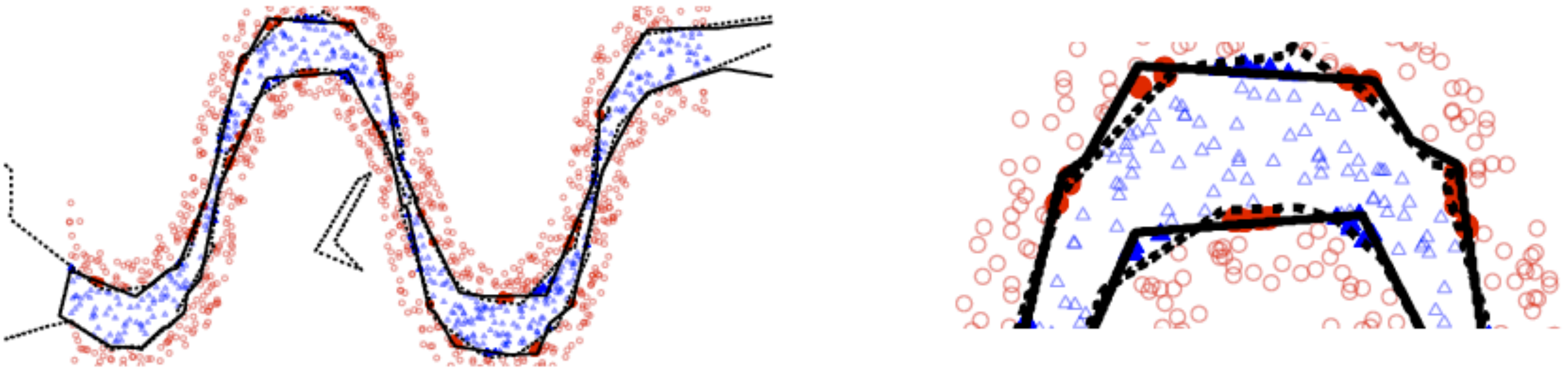


Figure 1: Binary classification using a shallow model with 20 hidden units (solid line) and a deep model with two layers of 10 units each (dashed line). The right panel shows a close-up of the left panel. Filled markers indicate errors made by the shallow model.

Established 1911

Deep Learning

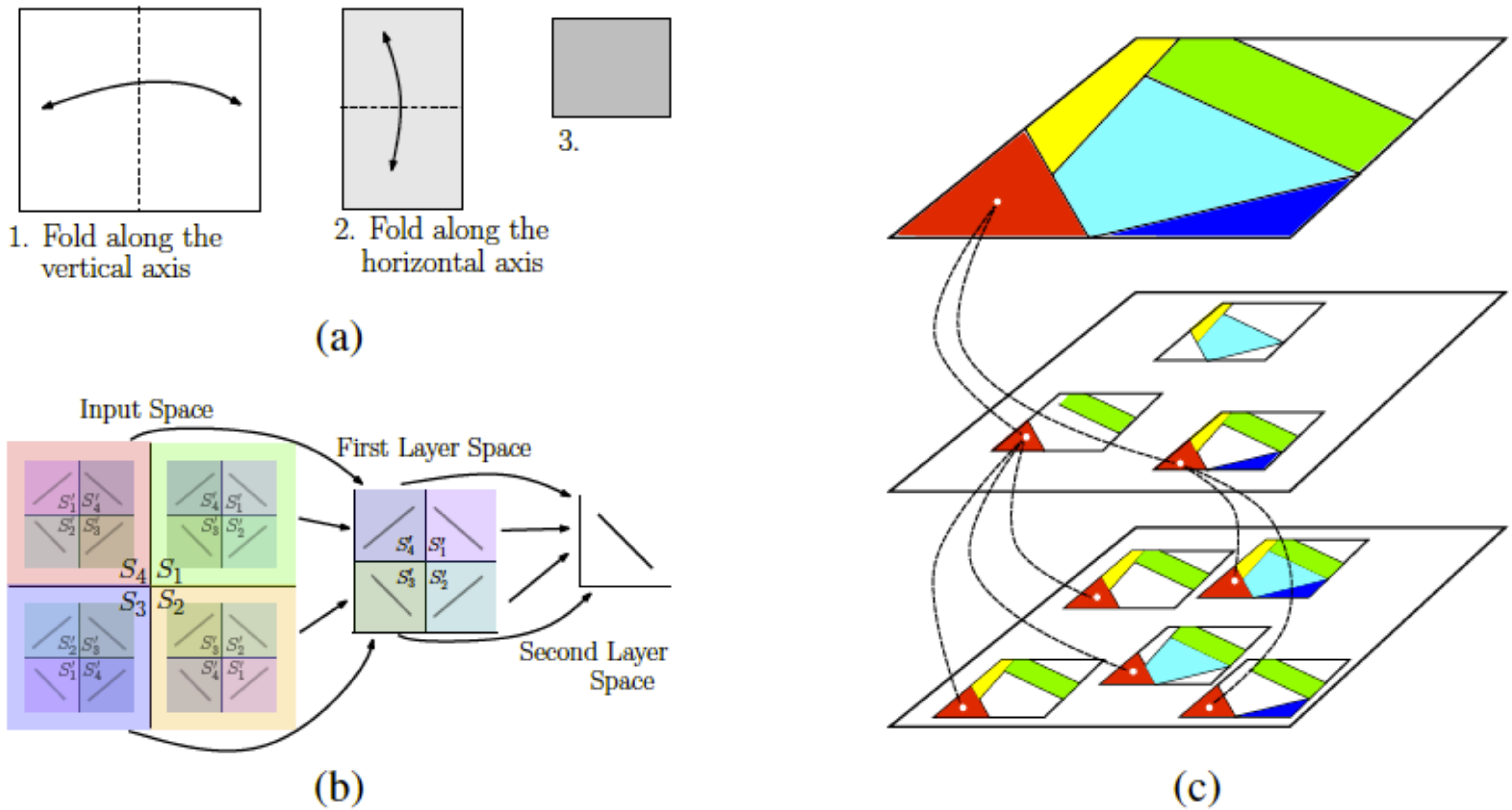


Figure 2: (a) Space folding of 2-D Euclidean space along the two axes. (b) An illustration of how the top-level partitioning (on the right) is replicated to the original input space (left). (c) Identification of regions across the layers of a deep model.

Deep Learning

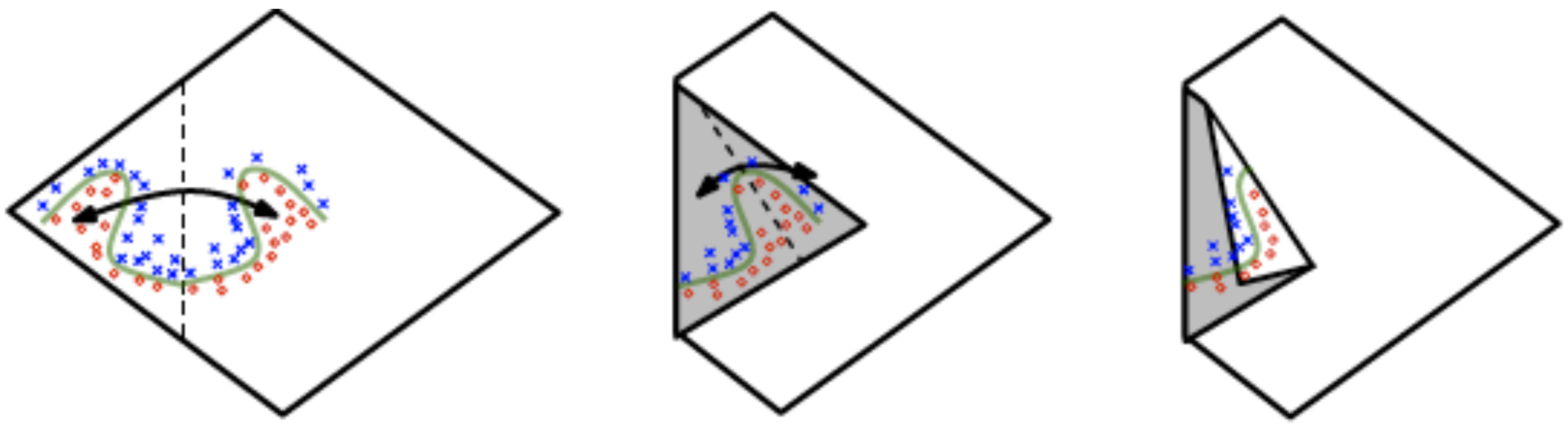


Figure 3: Space folding of 2-D space in a non-trivial way. Note how the folding can potentially identify symmetries in the boundary that it needs to learn.

Established 1911