Machine Learning for Sensory Signals

Neural Networks and Deep Learning

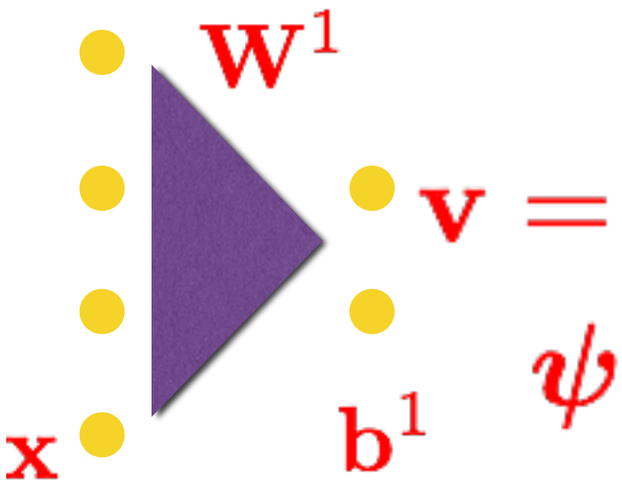
13-04-2017

Established 1911

INDIAN INSTITUTE OF SCIENCE
भारतीय विज्ञान संस्थान

Neural Networks

Perceptron Model [McCulloch, 1943, Rosenblatt, 1957]



$$\mathbf{v} = \psi(\mathbf{W}^1\mathbf{x} + \mathbf{b}^1)$$

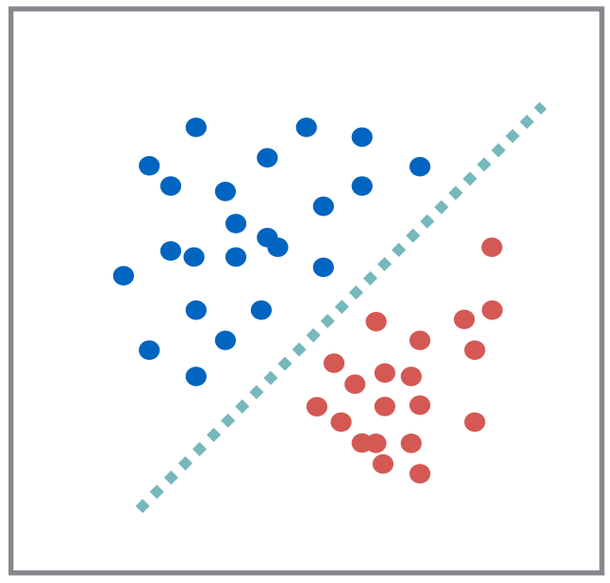
ψ thresholding function

Targets are binary classes [0,1]

Early Application for simulating the brain

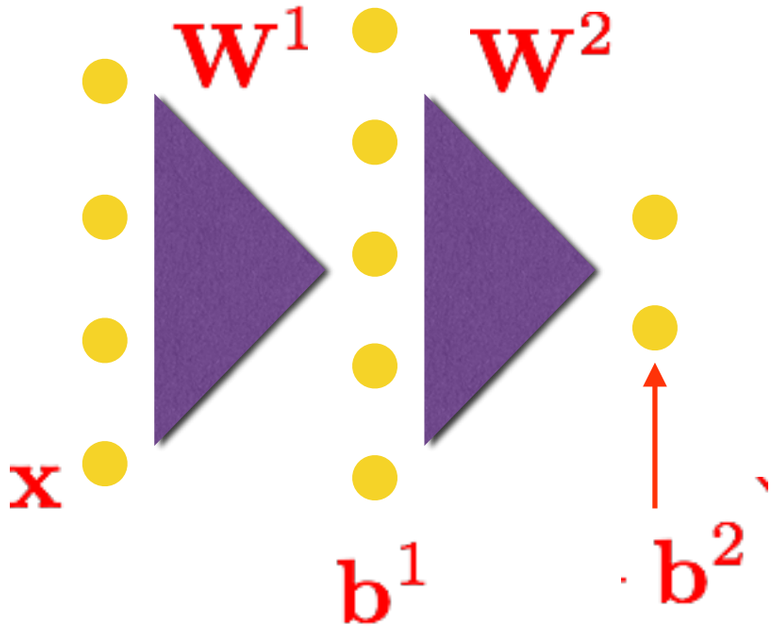
Data - not linearly separable often

Useful for linearly separable data



Neural Networks

Multi-layer Perceptron [Hopfield, 1982]



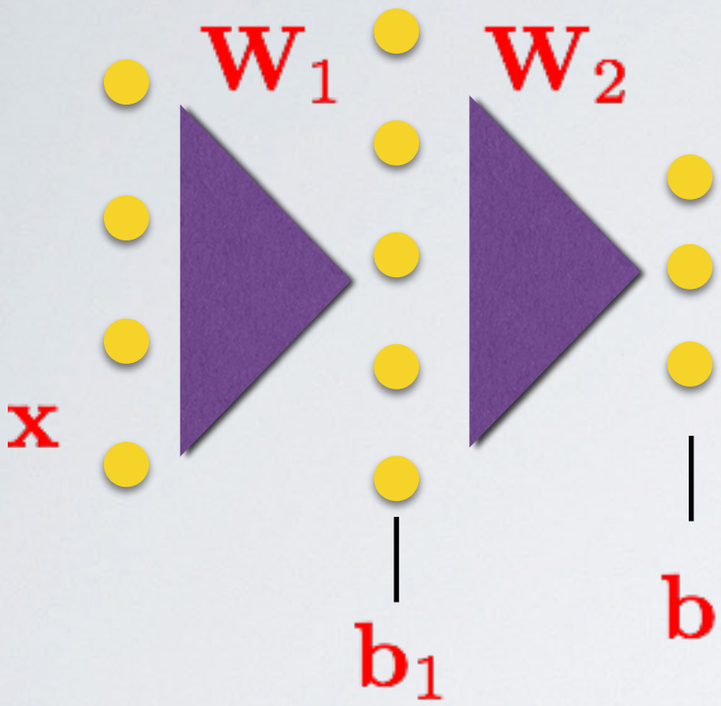
$$\mathbf{v}^2 = \psi\left(\mathbf{W}^2\phi(\mathbf{W}^1\mathbf{x} + \mathbf{b}^1) + \mathbf{b}^2\right)$$

ϕ non-linear function (*tanh*,*sigmoid*)

ψ thresholding function

- Useful for classifying non-linear data.
- Modeling of complex real world data like speech and images is some-what limited.

Parameter Learning

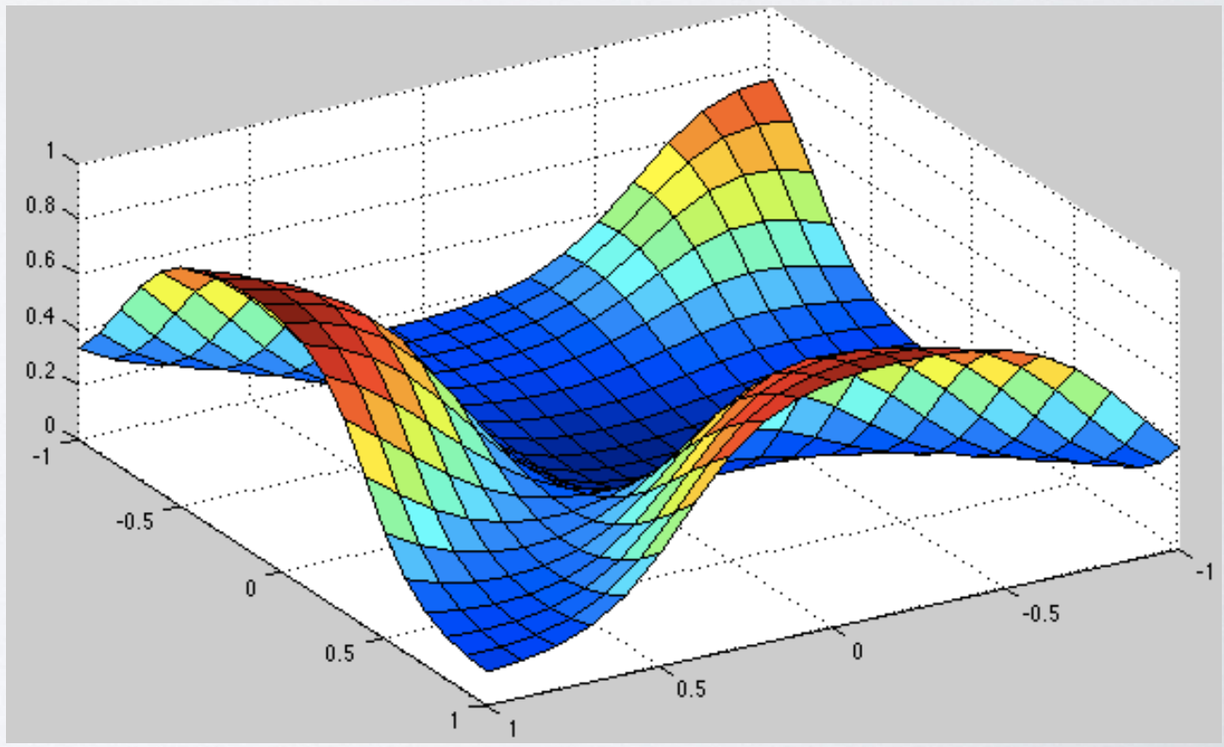


$$\mathbf{v}^2 = \psi\left(\mathbf{W}^2\phi(\mathbf{W}^1\mathbf{x} + \mathbf{b}^1) + \mathbf{b}^2\right)$$

Error function for entire data

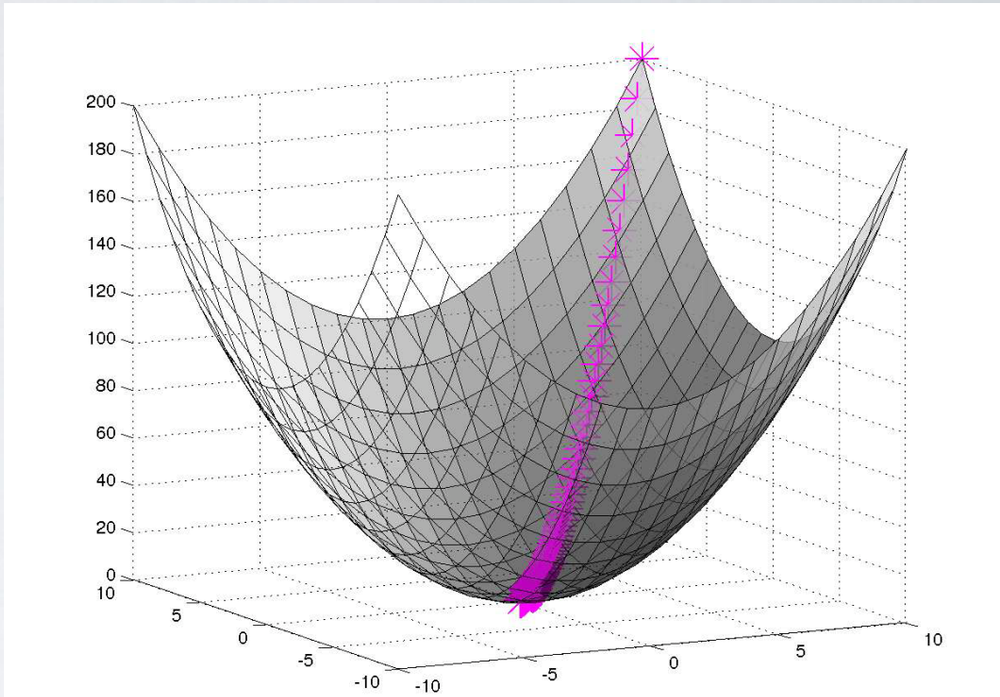
$$J_{MSE} = \sum_{i=1}^{M} ||\mathbf{v}_i - \mathbf{y}_i||^2$$

Typical Error Surface as a function of parameters (weights and biases)



Parameter Learning

Error surface close to a local optima



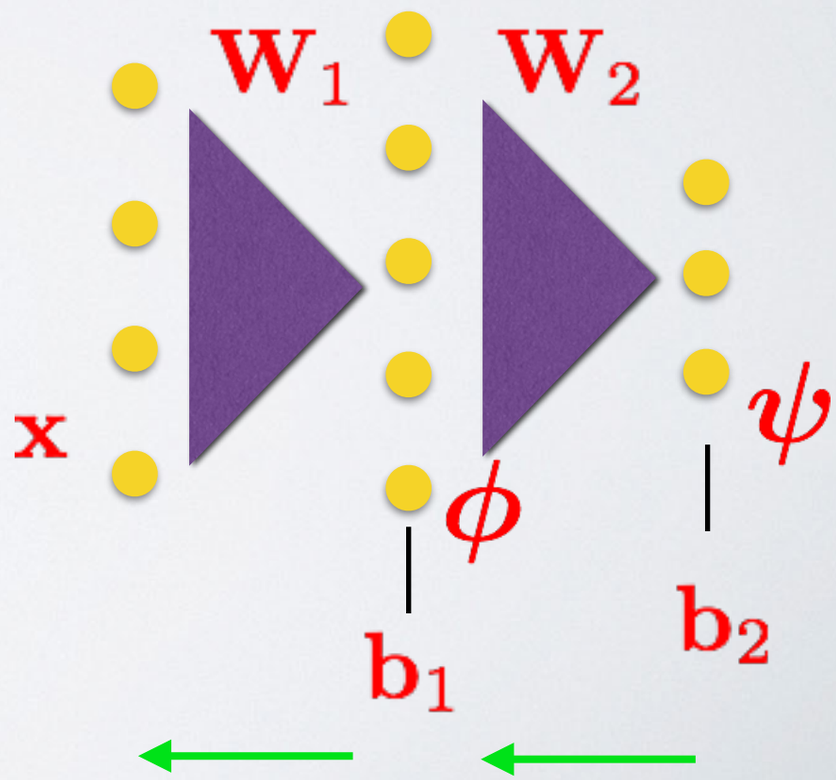
Non-linear nature of error function

- Move in the reverse direction of the gradient

$$\boldsymbol{W}_1^t = \boldsymbol{W}_1^{t-1} - \eta \frac{\partial J}{\partial \boldsymbol{W}_1}$$

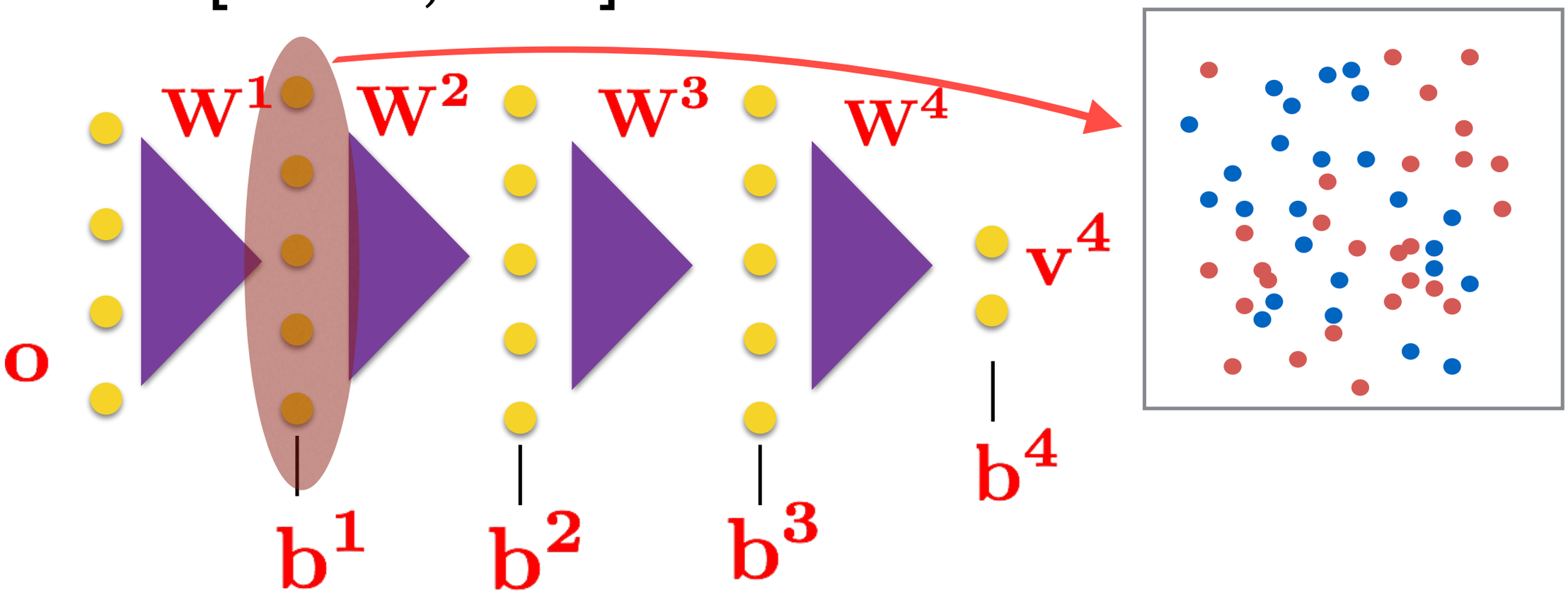
Error back propagation

$$\frac{\partial J}{\partial \mathbf{W}_1} = \frac{\partial J}{\partial \psi} \times \frac{\partial \psi}{\partial \phi} \times \frac{\partial \phi}{\partial \mathbf{W}_1}$$



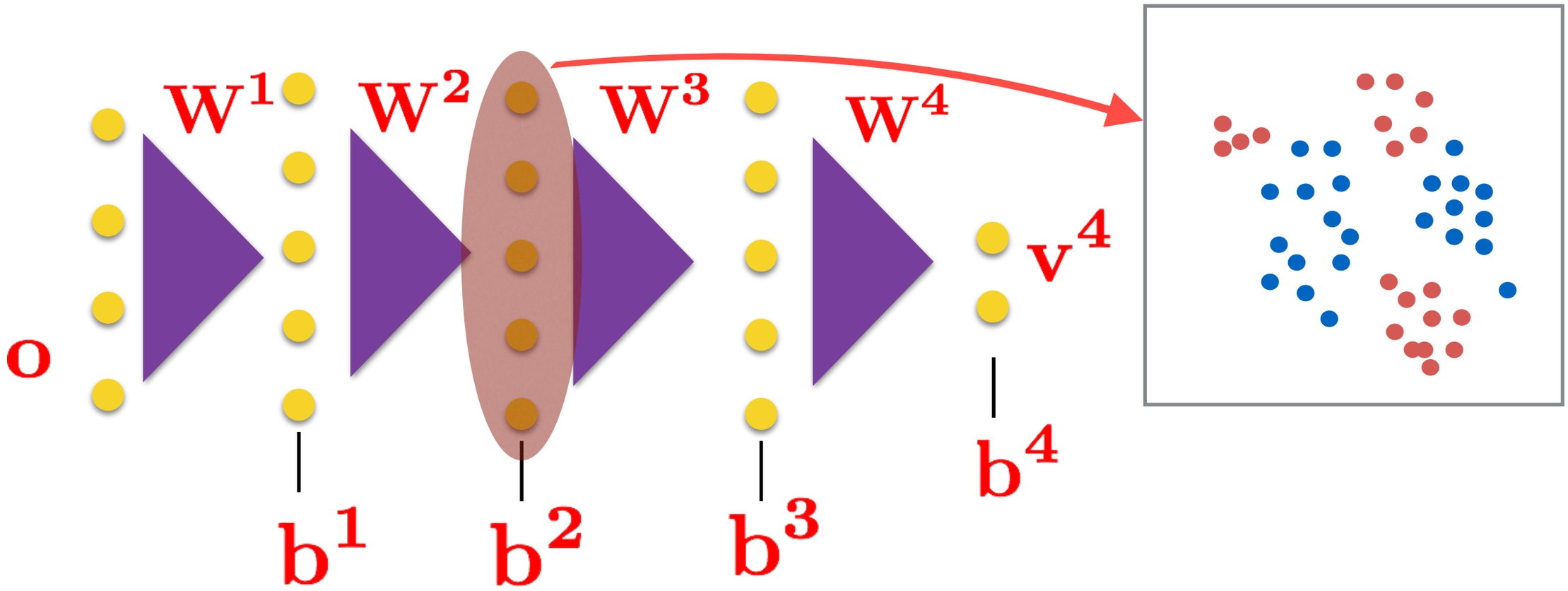
Deep Networks

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



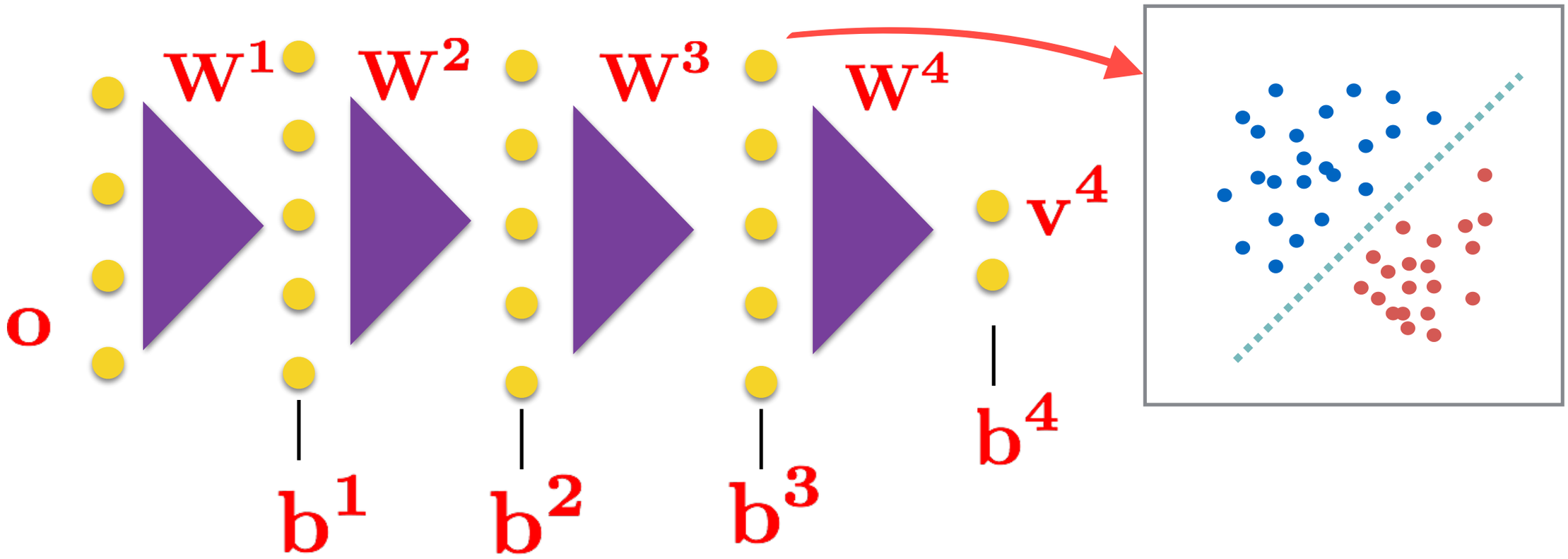
Deep Networks

Neural networks with multiple hidden layers - Deep networks



Deep Networks

Neural networks with multiple hidden layers - Deep networks



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