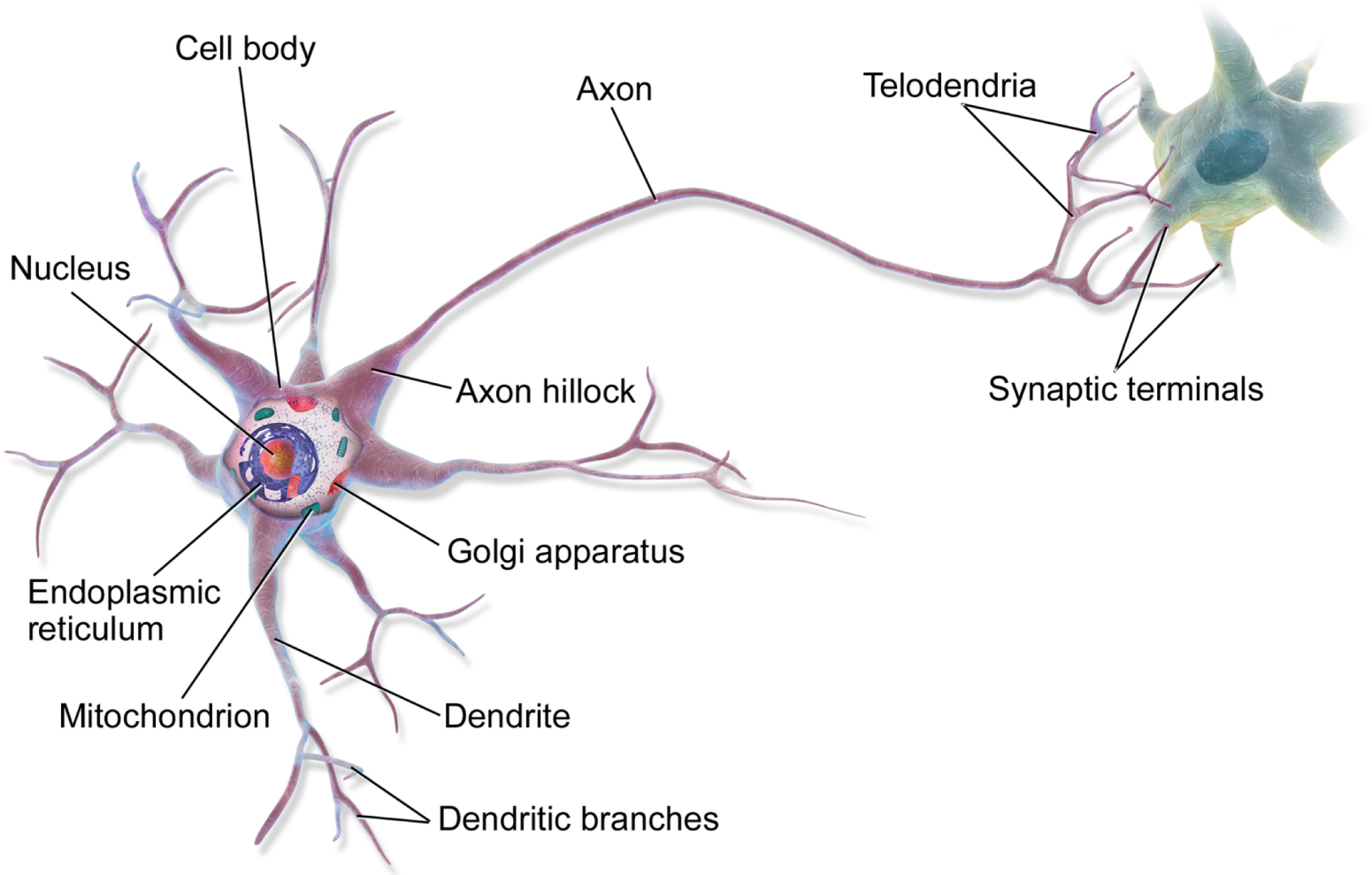
NEW YORK UNIVERSITY

DLI Teaching Kit

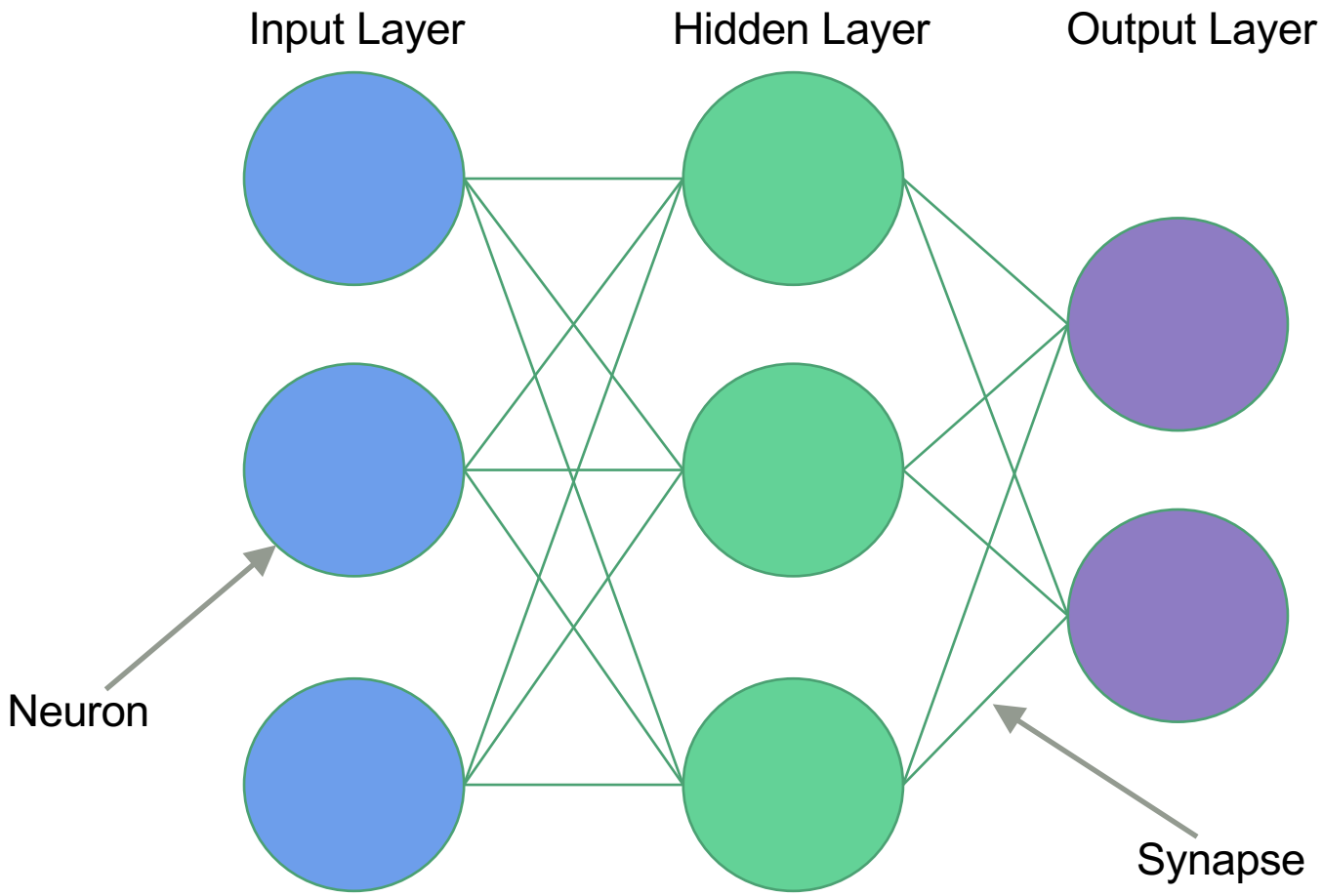
Lecture 1.3 - Introduction to Neural Networks

Deck credit: J. Seng

Biological Inspiration



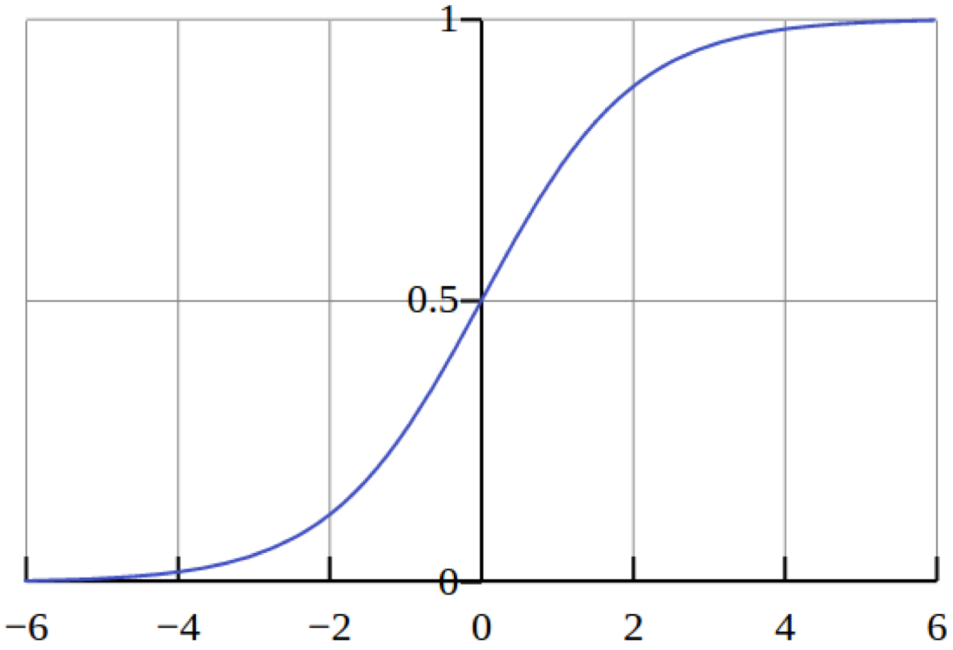
Neural Network Architecture



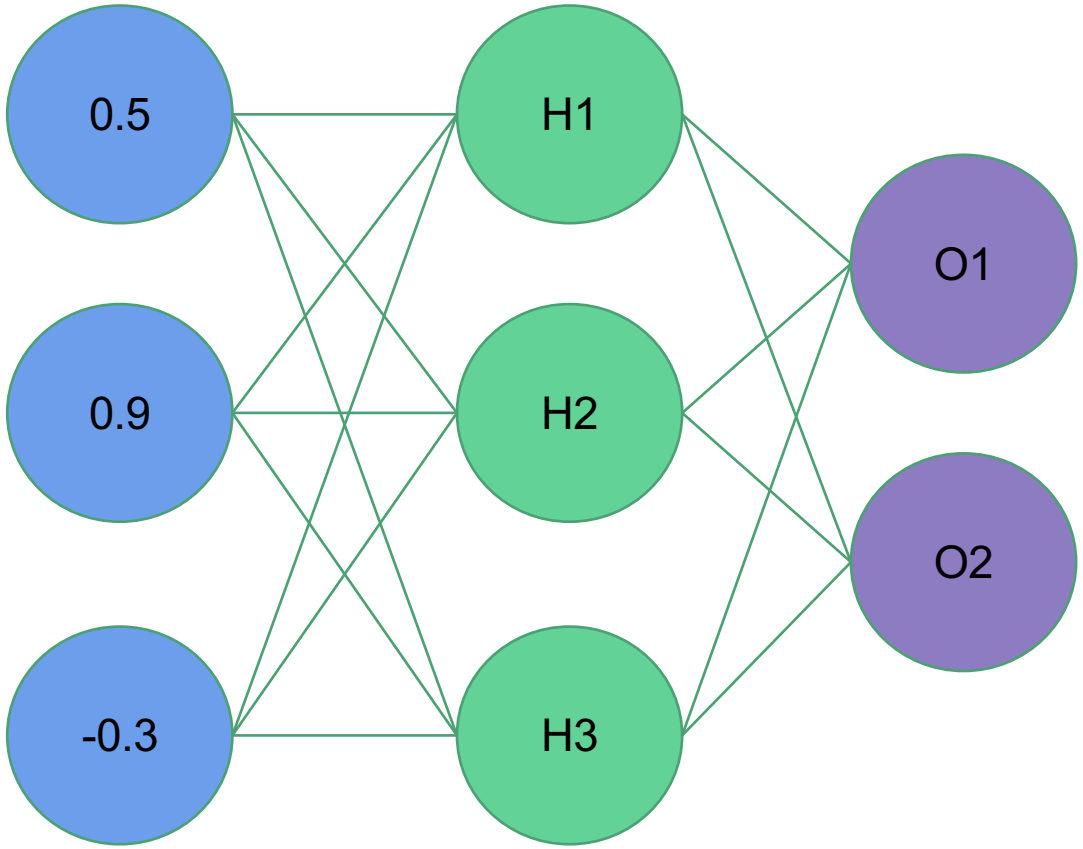
Activation Functions

- Activation Functions are applied to the inputs at each neuron
  - A common activation function is the Sigmoid

$$S(t) = \frac{1}{1 + e^{-t}}$$



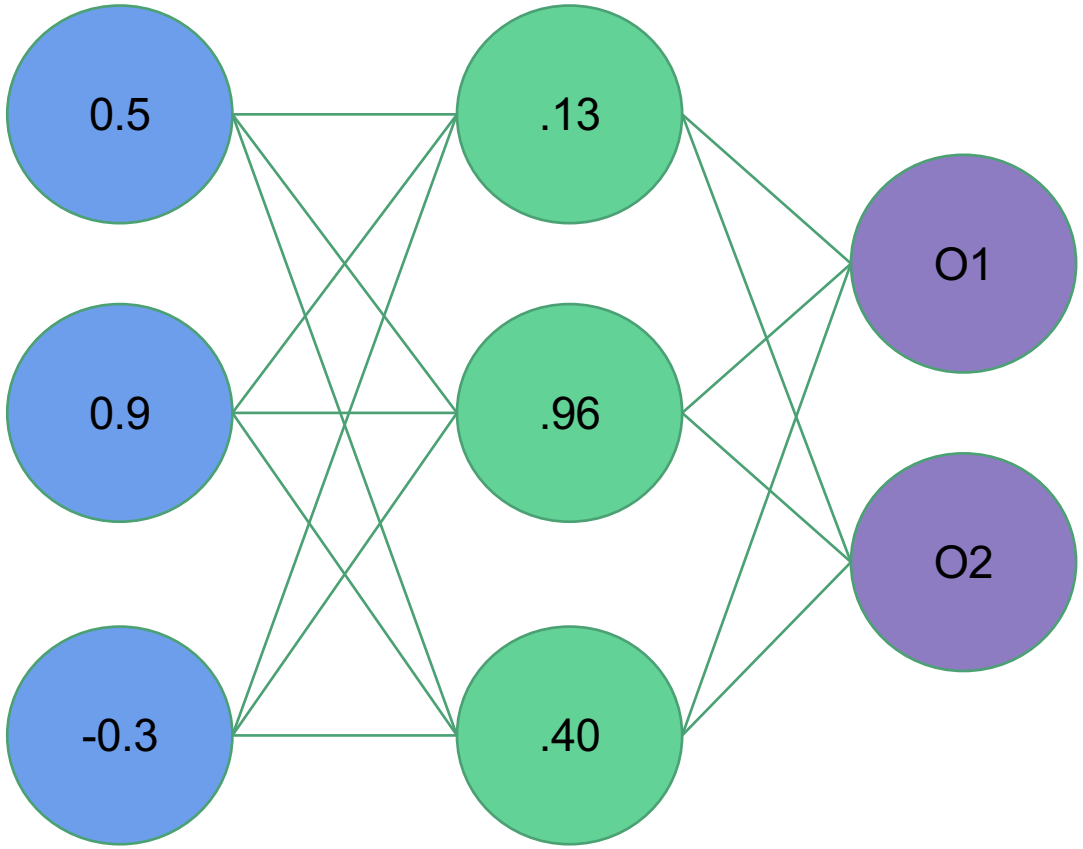
Inference



H1 Weights = (1.0, -2.0, 2.0)
H2 Weights = (2.0, 1.0, -4.0)
H3 Weights = (1.0, -1.0, 0.0)

O1 Weights = (-3.0, 1.0, -3.0)
O2 Weights = (0.0, 1.0, 2.0)

Inference

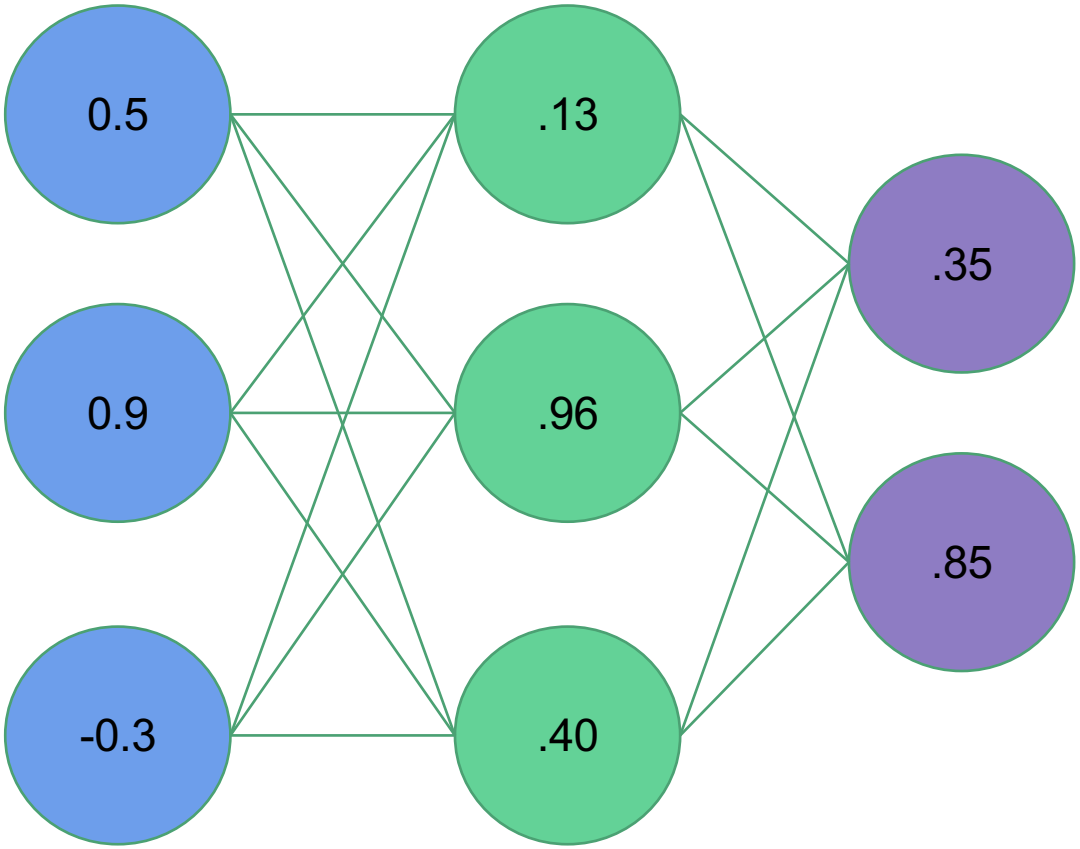


H1 Weights = (1.0, -2.0, 2.0)
H2 Weights = (2.0, 1.0, -4.0)
H3 Weights = (1.0, -1.0, 0.0)

O1 Weights = (-3.0, 1.0, -3.0)
O2 Weights = (0.0, 1.0, 2.0)

H1 = S(0.5 * 1.0 + 0.9 * -2.0 + -0.3 * 2.0) = S(-1.9) = .13
H2 = S(0.5 * 2.0 + 0.9 * 1.0 + -0.3 * -4.0) = S(3.1) = .96
H3 = S(0.5 * 1.0 + 0.9 * -1.0 + -0.3 * 0.0) = S(-0.4) = .40

Inference



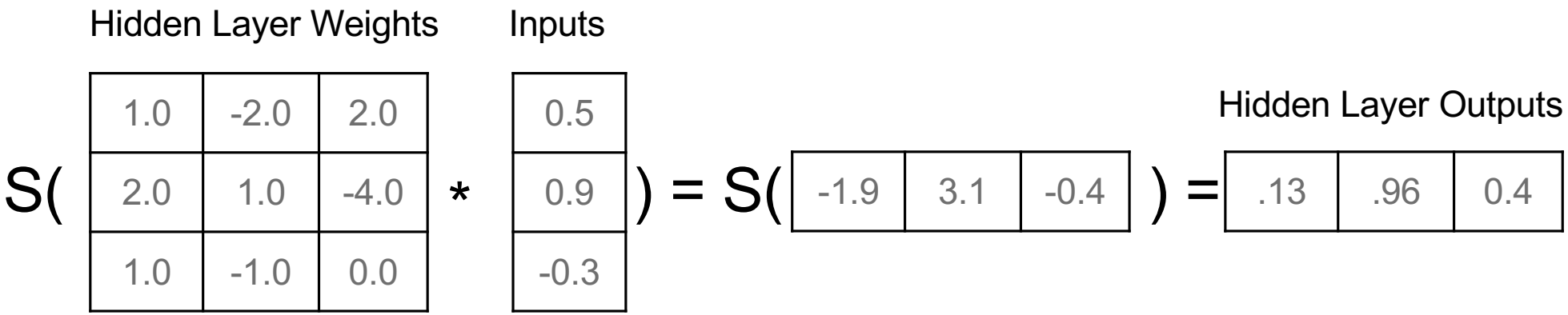
H1 Weights = (1.0, -2.0, 2.0)
H2 Weights = (2.0, 1.0, -4.0)
H3 Weights = (1.0, -1.0, 0.0)

O1 Weights = (-3.0, 1.0, -3.0)
O2 Weights = (0.0, 1.0, 2.0)

O1 = S(.13 * -3.0 + .96 * 1.0 + .40 * -3.0) = S(-.63) = .35
O1 = S(.13 * 0.0 + .96 * 1.0 + .40 * 2.0) = S(1.76) = .85

Matrix Formulation

H1 Weights = (1.0, -2.0, 2.0)
H2 Weights = (2.0, 1.0, -4.0)
H3 Weights = (1.0, -1.0, 0.0)



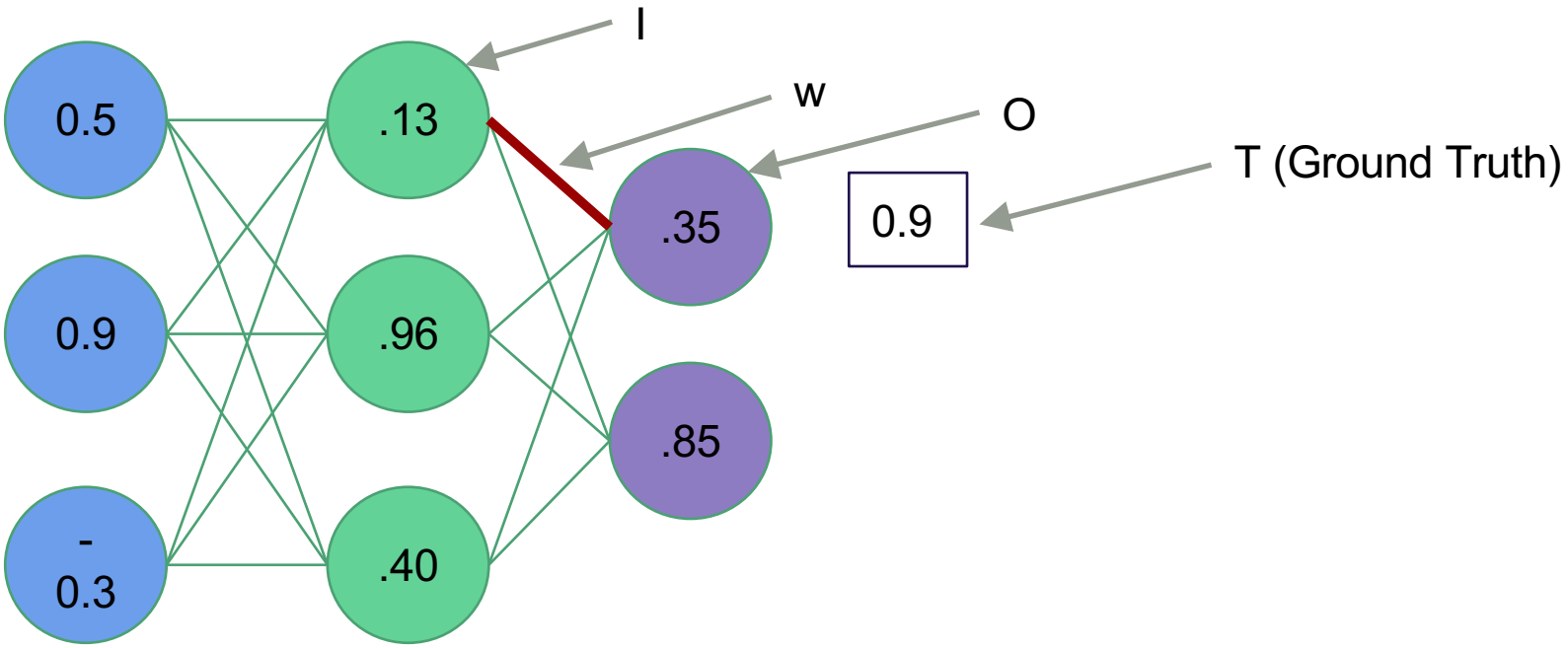
Training Neural Networks

- Procedure for training Neural Networks
  - Perform inference on the training set
  - Calculate the error between the predictions and actual labels of the training set
  - Determine the contribution of each Neuron to the error
  - Modify the weights of the Neural Network to minimize the error
- Error contributions are calculated using Backpropagation
- Error minimization is achieved with Gradient Descent

Backpropagation

- Problem: Which weights should be updated and by how much?
  - Insight: Use the derivative of the error with respect to weight to assign “blame”

Backpropagation Example



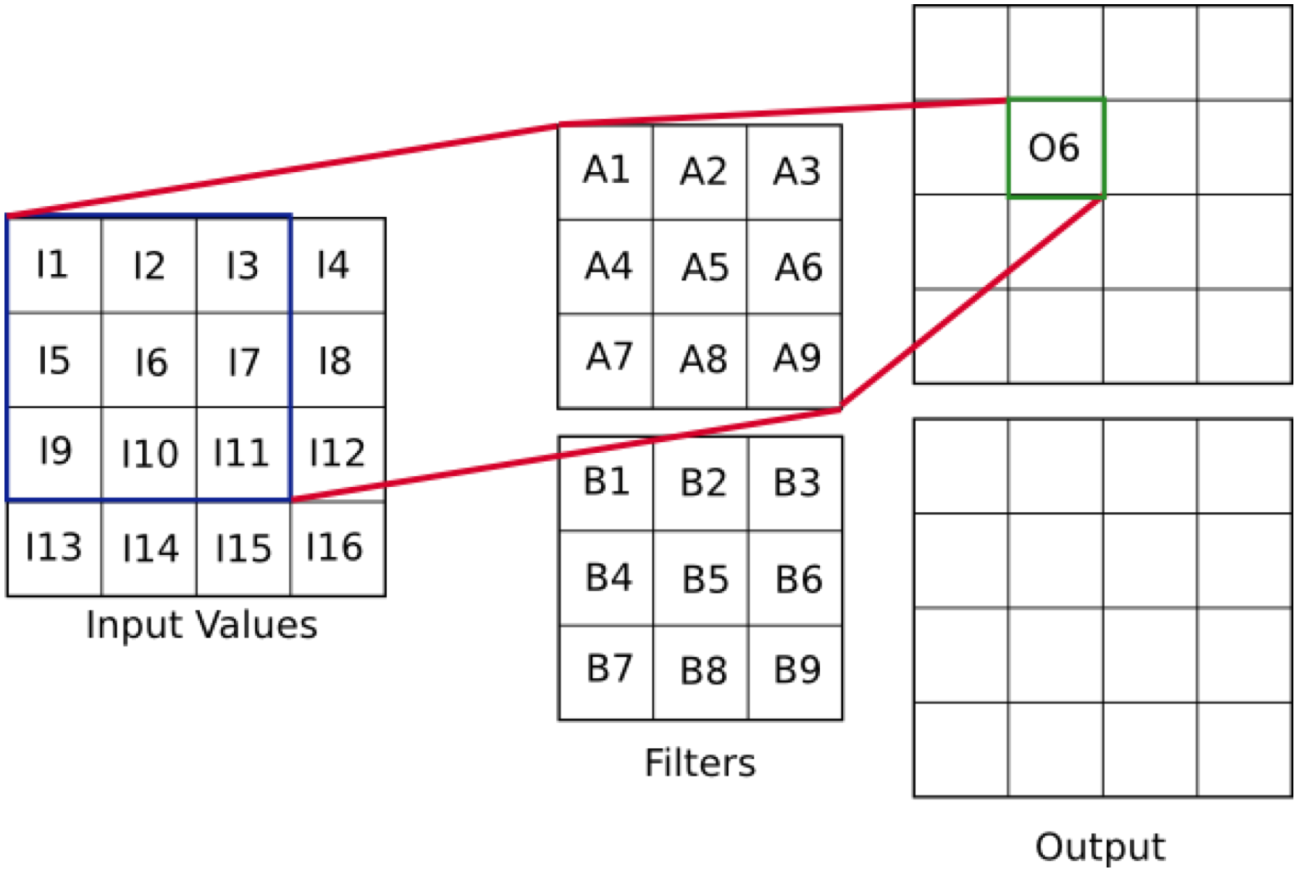
$$\frac{\partial E}{\partial w} = I \cdot (O - T) \cdot O \cdot (1 - O)$$

$$\frac{\partial E}{\partial w} = .13 \cdot (.35 - .9) \cdot .35 \cdot (1 - .35)$$

Gradient Descent

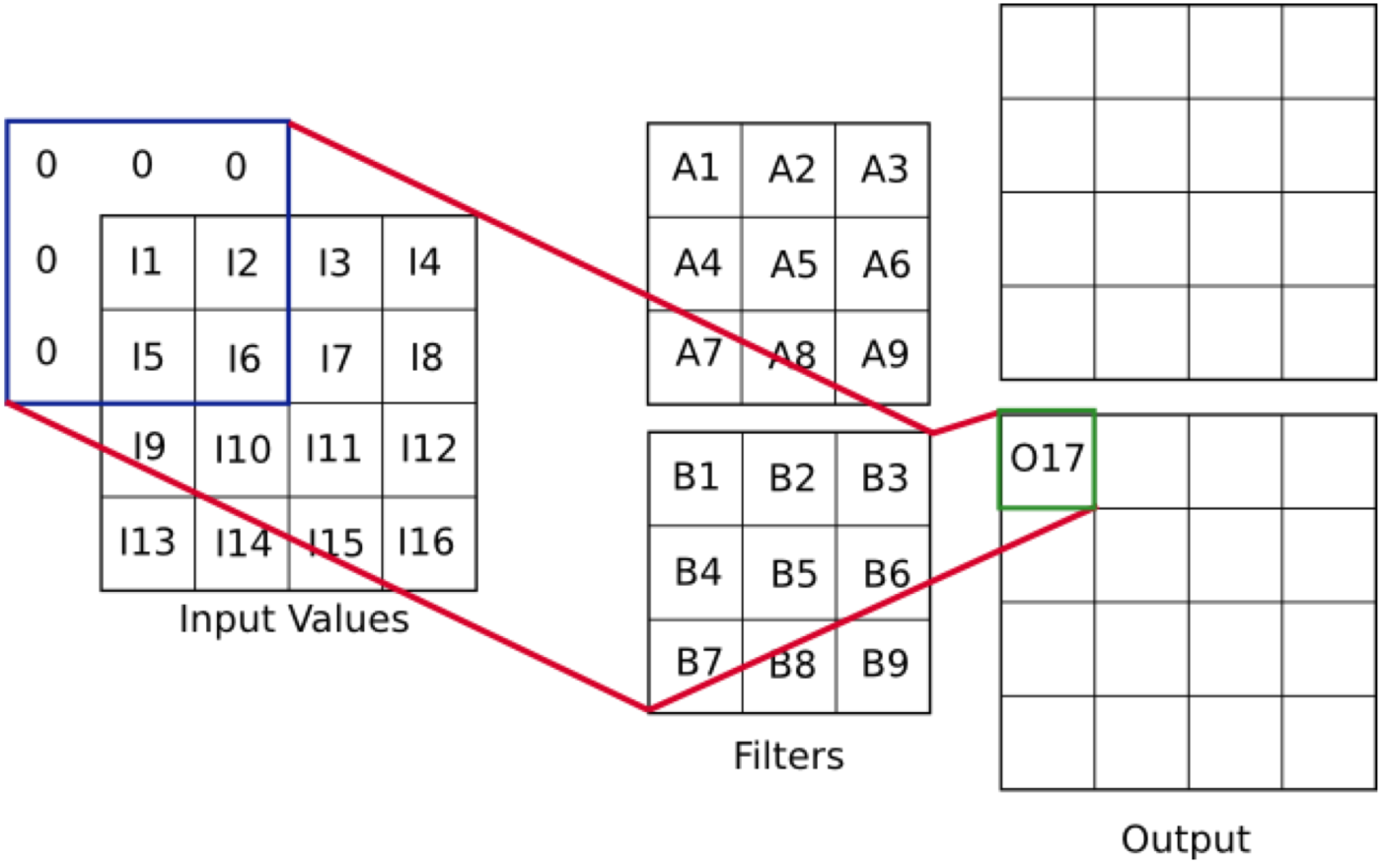
- Gradient Descent minimizes the neural network's error
  - At each time step the error of the network is calculated on the training data
  - Then the weights are modified to reduce the error
- Gradient Descent terminates when
  - The error is sufficiently small
  - The max number of time steps has been exceeded

Convolutional Neural Networks



$$O_6 = A_1 \cdot I_1 + A_2 \cdot I_2 + A_3 \cdot I_3 + A_4 \cdot I_5 + A_5 \cdot I_6 + A_6 \cdot I_7 + A_7 \cdot I_9 + A_8 \cdot I_{10} + A_9 \cdot I_{11}$$

Convolutional Neural Networks



$$O_{17} = B_5 \cdot I_1 + B_6 \cdot I_2 + B_8 \cdot I_5 + B_9 \cdot I_6$$

Max-Pooling

NEW YORK UNIVERSITY

DLI Teaching Kit

Thank you