CISC/CMPE452/COGS 400
# Multi-layer Perceptrons with Backpropagation

Ch. 3 Text book

Farhana Zulkernine

Multilayer Networks

- Although single-layer perceptron networks can distinguish between any number of classes, they require linear separability of inputs.

- To overcome this limitation, we can use multiple layers of neurons.

- Rosenblatt first suggested this idea in 1961, but he used perceptrons.

How MLP Works

- Suppose we added another layer whose inputs are the pixel values of the character image.
  - Middle layers "learn" to group features effectively in multiple iterations
    - *Grouping is not known*
  - Then top layer uses that for classification.
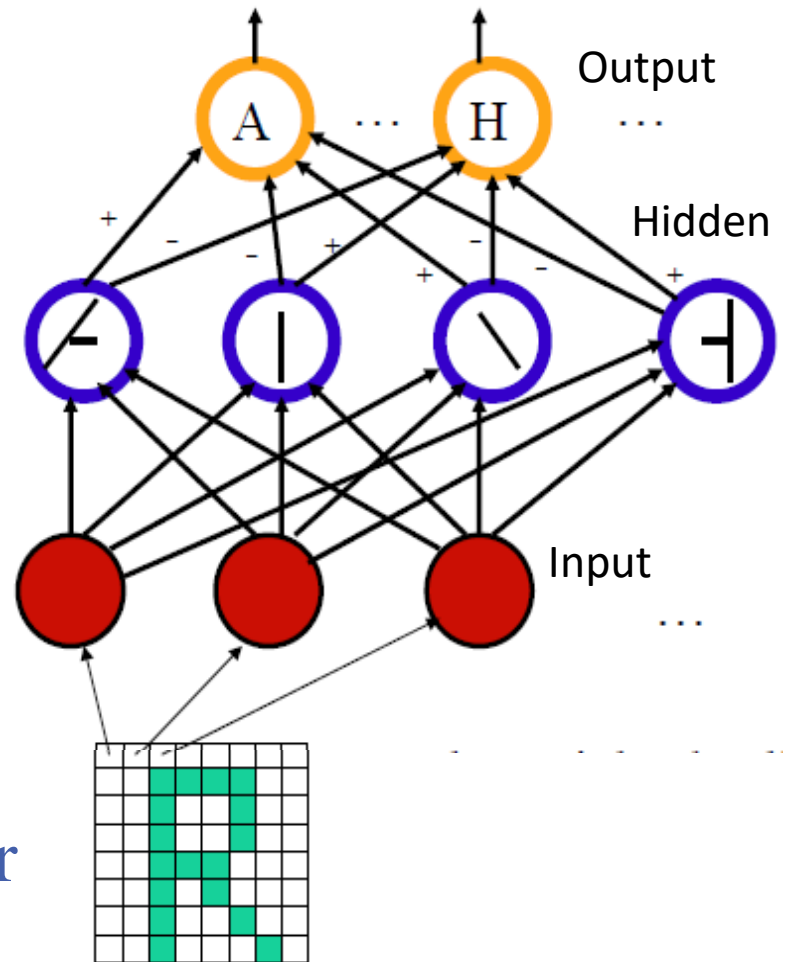
Problem with Training

- The weights leading to the top layer could be trained with examples, but in the 1960's there were *no algorithms for training the weights leading to the middle layer*.

- Example 2: Input may be a person's movement data from cell phone and top layer classifying the activity.
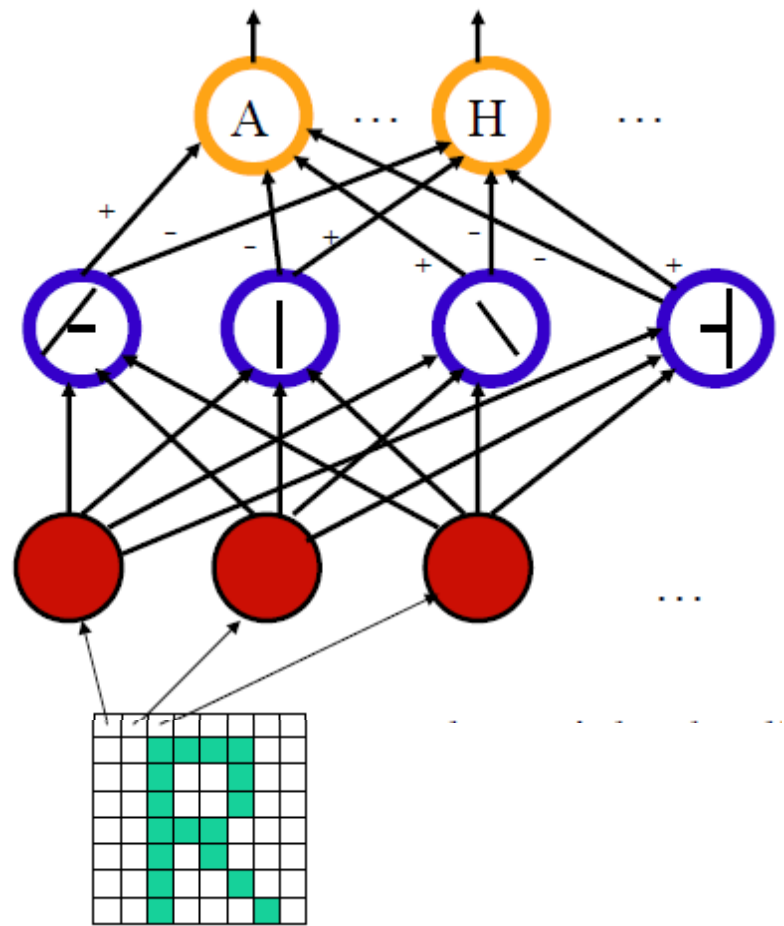
Activation and Loss Function

- The non-differentiable output function (e.g. step function gave 0, linear output gave constant value) led to an inefficient and weak learning algorithm. We could not compute how the actual output changed with weights - used only the change in error.

- The idea that eventually led to a breakthrough was the use of continuous output functions and gradient descent from Adaline.

- Here we use **sigmoid output/activation function** and **MSE with gradient descent**, both of which are differentiable.

Backpropagation Training Algorithm

- Multilayer networks introduced the **credit-assignment problem** – adjust weights based on their contribution to the error – how?

- The solution – **backpropagation** (BP) – which computes errors in a node and propagates the correction to the weight values backwards through the hidden nodes

- **Backpropagation algorithm** was popularized by Rumelhart, Hinton, and Williams (1986)

Supervised Learning and FFN

- BP assumes that the correct output (ground truth) is known to be able to compute the error.
  - Also known as supervised learning – supervisor knows the correct answer and helps reduce the error through multiple training iterations.
- The BP algorithm assumes a **feedforward** neural network architecture where nodes are partitioned into layers $\{0, L\}$.
  - The input layer is layer 0 and the output layer is L.
- BP addresses networks for which $L \geq 2$, containing "hidden layers" numbered 1 to $L - 1$.
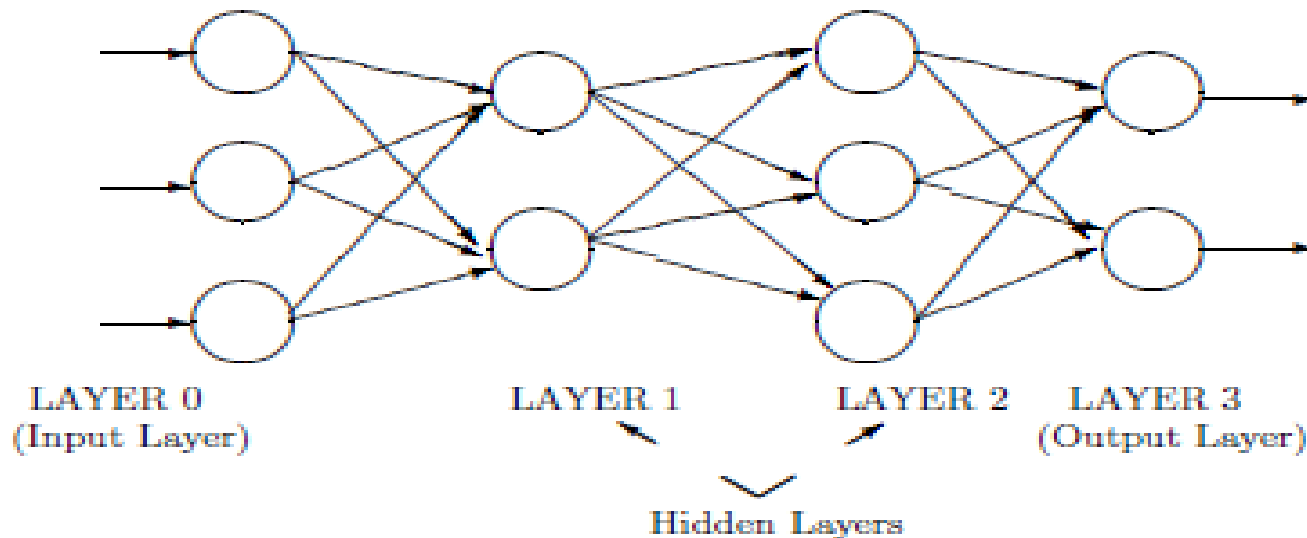
Feedforward Networks

Figure 1.12: Feed Froward 3-2-3-2 Network

- **Feedforward Networks:** A connection is allowed from a node in layer i only to nodes in layer i + 1.
  - Feedforward networks are succinctly described by a **sequence of numbers** indicating the number of nodes in each layer.
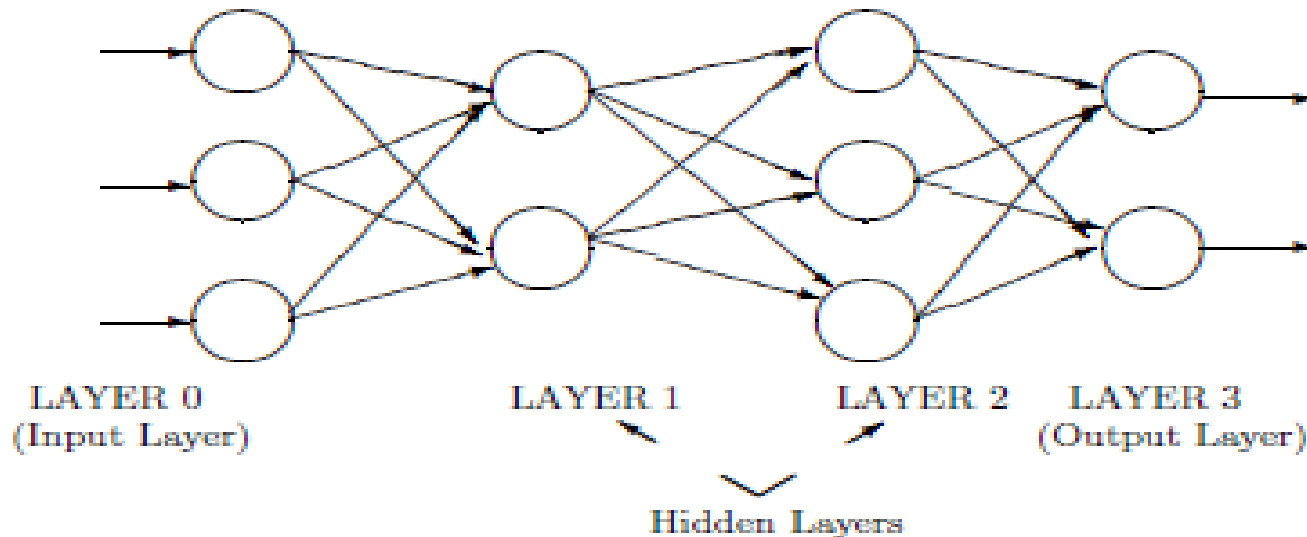
9

Feedforward Networks (cont…)

Figure 1.12: Feed Froward 3-2-3-2 Network

- **Feedforward Networks:** These networks, generally with *no more than 4 hidden layers*, are among the **most common** neural nets in use.
- Conceptually, nodes in successively higher layers abstract successively higher level features from preceding layers.
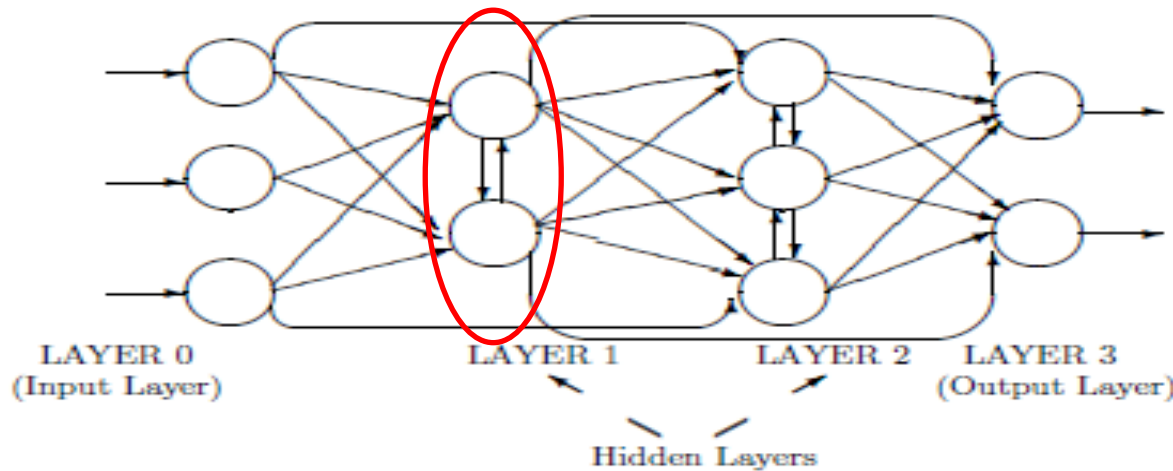
Layered Network

Figure 1.11: Layered Networks

- When nodes are partitioned or organized into subsets called **layers**, with connections only leading from layer k to layer j where j > k, no reverse connection from j to k.

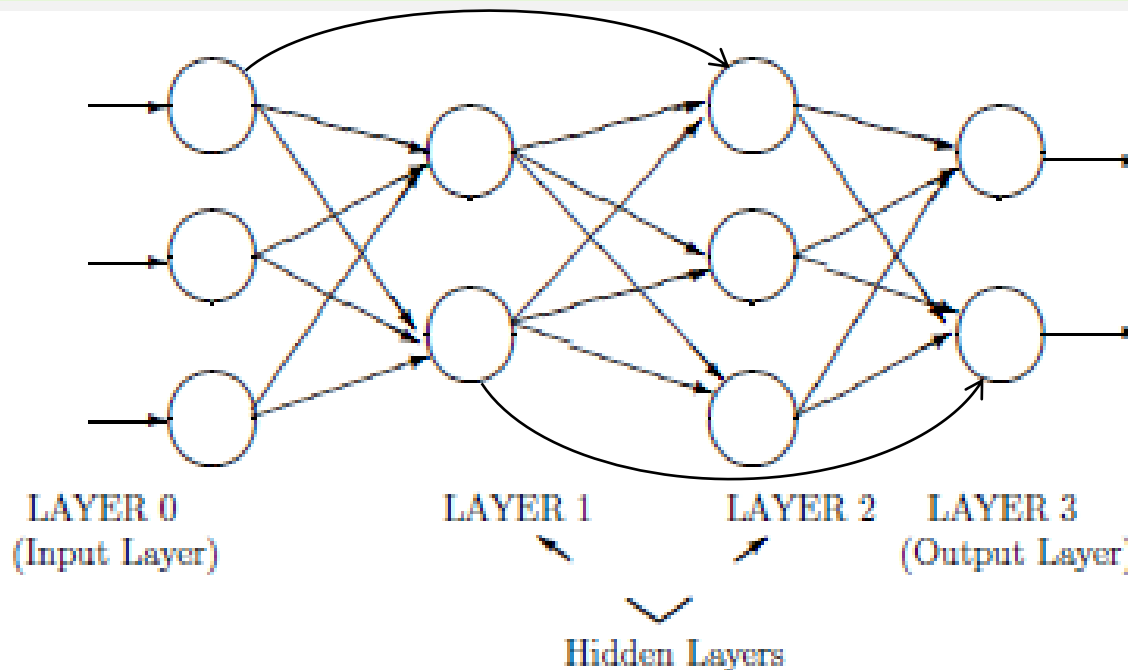- Intra-layer connections exist.

11

Acyclic Networks

Figure 1.12: Feed Froward 3-2-3-2 Network

- **Acyclic Networks:** Layered networks **without intra-layer connections**.