Week 9: Neural Network

Ekarat Rattagan

July 28, 2025

Motivation (1/2)

Given a dataset D,

x_1	x_2
0	1
2	3
4	5

How many coefficients θ_j are there for a polynomial of degree p?

Motivation (2/2)

How about this dataset,

Figure: A cat W100xH100 istockphoto 1361394182

How many coefficients θ_j are there for a polynomial of degree p?

Artificial Neural Network

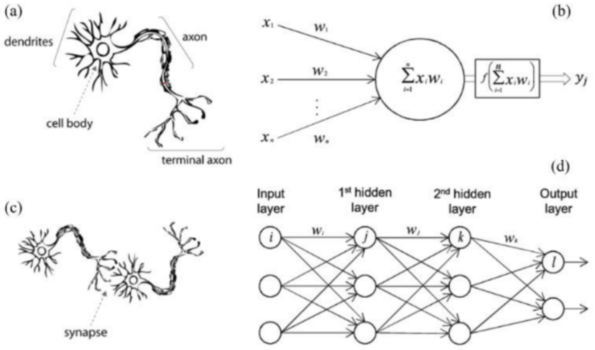


Figure 3. A biological neuron in comparison to an artificial neural network: (**a**) human neuron; (**b**) artificial neuron; (**c**) biological synapse; and (**d**) ANN synapses [39].

Figure: Artificial Neural Network

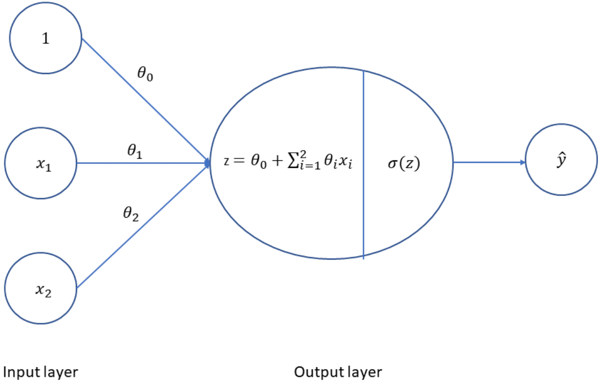
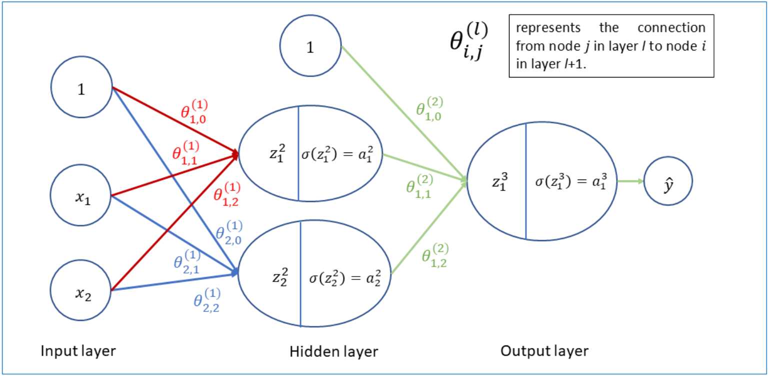


Figure: A simple Neural Network: Perceptron

Artificial Neural Network



$$a_1^2 = \sigma(\theta_{1,0}^{(1)} + \theta_{1,1}^{(1)} x_1 + \theta_{1,2}^{(1)} x_2)$$

$$a_2^2 = \sigma(\theta_{2,0}^{(1)} + \theta_{2,1}^{(1)} x_1 + \theta_{2,2}^{(1)} x_2)$$

$$a_1^3 = \sigma(\theta_{1,0}^{(2)} + \theta_{1,1}^{(2)} a_1^2 + \theta_{1,2}^{(2)} a_2^2)$$

$$\theta^{(1)} = \begin{bmatrix} \theta_{1,0}^{(1)} & \theta_{1,1}^{(1)} & \theta_{1,2}^{(1)} \\ \theta_{2,0}^{(1)} & \theta_{2,1}^{(1)} & \theta_{2,2}^{(1)} \end{bmatrix} \qquad \theta^{(2)} = \begin{bmatrix} \theta_{1,0}^{(2)} & \theta_{1,1}^{(2)} & \theta_{1,2}^{(2)} \end{bmatrix}$$

Figure: Multilayer Perceptron

Training Data Example

AND Operation

x_1	x_2	y
0	0	0
0	1	0
1	0	0
1	1	1

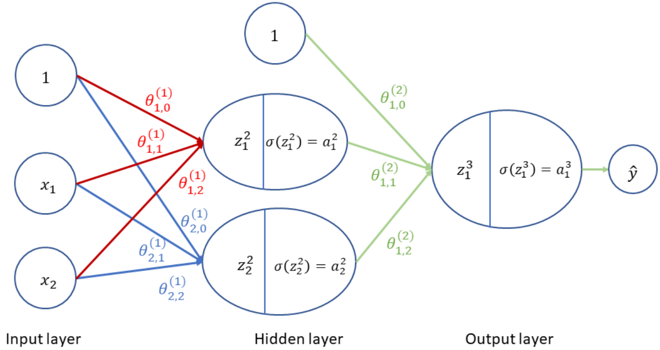
XNOR Operation

x_1	x_2	y
0	0	1
0	1	0
1	0	0
1	1	1

Training Process

- **Cost Function**: $J(\theta) = -y \log(\hat{y}) - (1-y)\log(1-\hat{y})$
- **Goal**: $\theta^* = \arg\min J(\theta)$
- **For each sample in the training data:**
  - Step 1: Forward pass
    Calculate $\hat{y}$, i.e. a^L, where a big L is the output layer.
  - Step 2: Calculate loss of the output layer, i.e. $a^L - y$.
  - Step 3: Backward pass (Backpropagation)
    Find $\frac{\partial J(\theta)}{\partial \theta_{ij}}$
  - Step 4: Update each $\theta_{i,j}^{(l)}$ (weights)

Step 1&2: Forward pass & Error of last layer



Step 3: Backward pass (Backpropagation) 1/6

$\theta^{(2)}$

- $\dfrac{\partial J(\theta)}{\partial \theta^{(2)}_{1,0}} = \dfrac{\partial J(\theta)}{\partial a^3_1} \cdot \dfrac{\partial a^3_1}{\partial z^3_1} \cdot \dfrac{\partial z^3_1}{\partial \theta^{(2)}_{1,0}}$

- $\dfrac{\partial J(\theta)}{\partial \theta^{(2)}_{1,1}} = \dfrac{\partial J(\theta)}{\partial a^3_1} \cdot \dfrac{\partial a^3_1}{\partial z^3_1} \cdot \dfrac{\partial z^3_1}{\partial \theta^{(2)}_{1,1}}$

- $\dfrac{\partial J(\theta)}{\partial \theta^{(2)}_{1,2}} = \dfrac{\partial J(\theta)}{\partial a^3_1} \cdot \dfrac{\partial a^3_1}{\partial z^3_1} \cdot \dfrac{\partial z^3_1}{\partial \theta^{(2)}_{1,2}}$

Step 3: Backward pass (Backpropagation) 2/6

$\theta^{(2)}$, **e.g.,** $\theta^{(2)}_{1,0}$

- $\frac{\partial J(\theta)}{\partial \theta^{(2)}_{1,0}} = \frac{\partial J(\theta)}{\partial a^3_1} \cdot \frac{\partial a^3_1}{\partial z^3_1} \cdot \frac{\partial z^3_1}{\partial \theta^{(2)}_{1,0}}$

- $$\begin{aligned}\frac{\partial J(\theta)}{\partial a^3_1} &= \frac{\partial(-y\log(a^3_1)-(1-y)\log(1-a^3_1))}{\partial a^3_1} \\ &= -\frac{y}{a^3_1} + \frac{1-y}{1-a^3_1}\end{aligned}$$

- $$\begin{aligned}\frac{\partial a^3_1}{\partial z^3_1} &= \frac{\partial \sigma(z^3_1)}{\partial z^3_1} \\ &= \sigma(z^3_1) \cdot (1 - \sigma(z^3_1)) \\ &= a^3_1 \cdot (1 - a^3_1)\end{aligned}$$

- $$\begin{aligned}\frac{\partial z^3_1}{\partial \theta^{(2)}_{1,0}} &= \frac{\partial(\theta^{(2)}_{1,0}+\theta^{(2)}_{1,1}.a^2_1+\theta^{(2)}_{1,2}.a^2_2)}{\partial \theta^{(2)}_{1,0}} \\ &= 1\end{aligned}$$

Step 3: Backward pass (Backpropagation) 3/6

$\delta^{(2)}$

- $\frac{\partial J(\theta)}{\partial \theta^{(2)}_{1,0}} = \frac{\partial J(\theta)}{\partial a^3_1} \cdot \frac{\partial a^3_1}{\partial z^3_1} \cdot \frac{\partial z^3_1}{\partial \theta^{(2)}_{1,0}} = (a^3_1 - y) \cdot 1$

- $\frac{\partial J(\theta)}{\partial \theta^{(2)}_{1,1}} = \frac{\partial J(\theta)}{\partial a^3_1} \cdot \frac{\partial a^3_1}{\partial z^3_1} \cdot \frac{\partial z^3_1}{\partial \theta^{(2)}_{1,1}} = (a^3_1 - y) \cdot a^2_1$

- $\frac{\partial J(\theta)}{\partial \theta^{(2)}_{1,2}} = \frac{\partial J(\theta)}{\partial a^3_1} \cdot \frac{\partial a^3_1}{\partial z^3_1} \cdot \frac{\partial z^3_1}{\partial \theta^{(2)}_{1,2}} = (a^3_1 - y) \cdot a^2_2$

- Let define $\delta^3_1 = (a^3_1 - y)$ be the error of the first node in the last layer ($L = 3$ in this example).

- We then get, $\frac{\partial J(\theta)}{\partial \theta^{(l)}_{i,j}} = \delta^{l+1}_i \cdot a^l_j$

Step 3: Backward pass (Backpropagation) 4/6

$\theta^{(1)}$

- $\dfrac{\partial J(\theta)}{\partial \theta^{(1)}_{1,0}} = \dfrac{\partial J(\theta)}{\partial a^3_1} \cdot \dfrac{\partial a^3_1}{\partial z^3_1} \cdot \dfrac{\partial z^3_1}{\partial a^2_1} \cdot \dfrac{\partial a^2_1}{\partial z^2_1} \cdot \dfrac{\partial z^2_1}{\partial \theta^{(1)}_{1,0}}$
- $\dfrac{\partial J(\theta)}{\partial \theta^{(1)}_{1,1}} = \dfrac{\partial J(\theta)}{\partial a^3_1} \cdot \dfrac{\partial a^3_1}{\partial z^3_1} \cdot \dfrac{\partial z^3_1}{\partial a^2_1} \cdot \dfrac{\partial a^2_1}{\partial z^2_1} \cdot \dfrac{\partial z^2_1}{\partial \theta^{(1)}_{1,1}}$
- $\dfrac{\partial J(\theta)}{\partial \theta^{(1)}_{1,2}} = \dfrac{\partial J(\theta)}{\partial a^3_1} \cdot \dfrac{\partial a^3_1}{\partial z^3_1} \cdot \dfrac{\partial z^3_1}{\partial a^2_1} \cdot \dfrac{\partial a^2_1}{\partial z^2_1} \cdot \dfrac{\partial z^2_1}{\partial \theta^{(1)}_{1,2}}$
- $\dfrac{\partial J(\theta)}{\partial \theta^{(1)}_{2,0}} = \dfrac{\partial J(\theta)}{\partial a^3_1} \cdot \dfrac{\partial a^3_1}{\partial z^3_1} \cdot \dfrac{\partial z^3_1}{\partial a^2_2} \cdot \dfrac{\partial a^2_2}{\partial z^2_2} \cdot \dfrac{\partial z^2_2}{\partial \theta^{(1)}_{2,0}}$
- $\dfrac{\partial J(\theta)}{\partial \theta^{(1)}_{2,1}} = \dfrac{\partial J(\theta)}{\partial a^3_1} \cdot \dfrac{\partial a^3_1}{\partial z^3_1} \cdot \dfrac{\partial z^3_1}{\partial a^2_2} \cdot \dfrac{\partial a^2_2}{\partial z^2_2} \cdot \dfrac{\partial z^2_2}{\partial \theta^{(1)}_{2,1}}$
- $\dfrac{\partial J(\theta)}{\partial \theta^{(1)}_{2,2}} = \dfrac{\partial J(\theta)}{\partial a^3_1} \cdot \dfrac{\partial a^3_1}{\partial z^3_1} \cdot \dfrac{\partial z^3_1}{\partial a^2_2} \cdot \dfrac{\partial a^2_2}{\partial z^2_2} \cdot \dfrac{\partial z^2_2}{\partial \theta^{(1)}_{2,2}}$

Step 3: Backward pass (Backpropagation) 5/6

$\theta^{(1)}$

- $\dfrac{\partial J(\theta)}{\partial \theta^{(1)}_{1,0}} = \dfrac{\partial J(\theta)}{\partial a^3_1} \cdot \dfrac{\partial a^3_1}{\partial z^3_1} \cdot \dfrac{\partial z^3_1}{\partial a^2_1} \cdot \dfrac{\partial a^2_1}{\partial z^2_1} \cdot \dfrac{\partial z^2_1}{\partial \theta^{(1)}_{1,0}} = (a^3_1 - y) \cdot \theta^{(2)}_{1,1} \cdot a^2_1(1 - a^2_1) \cdot 1$
- $\dfrac{\partial J(\theta)}{\partial \theta^{(1)}_{1,1}} = \dfrac{\partial J(\theta)}{\partial a^3_1} \cdot \dfrac{\partial a^3_1}{\partial z^3_1} \cdot \dfrac{\partial z^3_1}{\partial a^2_1} \cdot \dfrac{\partial a^2_1}{\partial z^2_1} \cdot \dfrac{\partial z^2_1}{\partial \theta^{(1)}_{1,1}} = (a^3_1 - y) \cdot \theta^{(2)}_{1,1} \cdot a^2_1(1 - a^2_1) \cdot a^1_1$
- $\dfrac{\partial J(\theta)}{\partial \theta^{(1)}_{1,2}} = \dfrac{\partial J(\theta)}{\partial a^3_1} \cdot \dfrac{\partial a^3_1}{\partial z^3_1} \cdot \dfrac{\partial z^3_1}{\partial a^2_1} \cdot \dfrac{\partial a^2_1}{\partial z^2_1} \cdot \dfrac{\partial z^2_1}{\partial \theta^{(1)}_{1,2}} = (a^3_1 - y) \cdot \theta^{(2)}_{1,1} \cdot a^2_1(1 - a^2_1) \cdot a^1_2$
- $\dfrac{\partial J(\theta)}{\partial \theta^{(1)}_{2,0}} = \dfrac{\partial J(\theta)}{\partial a^3_1} \cdot \dfrac{\partial a^3_1}{\partial z^3_1} \cdot \dfrac{\partial z^3_1}{\partial a^2_2} \cdot \dfrac{\partial a^2_2}{\partial z^2_2} \cdot \dfrac{\partial z^2_2}{\partial \theta^{(1)}_{2,0}} = (a^3_1 - y) \cdot \theta^{(2)}_{1,2} \cdot a^2_2(1 - a^2_2) \cdot 1$
- $\dfrac{\partial J(\theta)}{\partial \theta^{(1)}_{2,1}} = \dfrac{\partial J(\theta)}{\partial a^3_1} \cdot \dfrac{\partial a^3_1}{\partial z^3_1} \cdot \dfrac{\partial z^3_1}{\partial a^2_2} \cdot \dfrac{\partial a^2_2}{\partial z^2_2} \cdot \dfrac{\partial z^2_2}{\partial \theta^{(1)}_{2,1}} = (a^3_1 - y) \cdot \theta^{(2)}_{1,2} \cdot a^2_2(1 - a^2_2) \cdot a^1_1$
- $\dfrac{\partial J(\theta)}{\partial \theta^{(1)}_{2,2}} = \dfrac{\partial J(\theta)}{\partial a^3_1} \cdot \dfrac{\partial a^3_1}{\partial z^3_1} \cdot \dfrac{\partial z^3_1}{\partial a^2_2} \cdot \dfrac{\partial a^2_2}{\partial z^2_2} \cdot \dfrac{\partial z^2_2}{\partial \theta^{(1)}_{2,2}} = (a^3_1 - y) \cdot \theta^{(2)}_{1,2} \cdot a^2_2(1 - a^2_2) \cdot a^1_2$

Note that a^1_1 *is* x_1, a^1_2 *is* x_2

Step 3: Backward pass (Backpropagation) 6/6

Based on the observation above, we can conclude that

- $\delta_1^2 = \delta_1^3 \cdot \theta_{1,1}^{(2)} \cdot a_1^2(1 - a_1^2)$
- $\delta_2^2 = \delta_1^3 \cdot \theta_{1,2}^{(2)} \cdot a_2^2(1 - a_2^2)$
- let define $\delta_i^l = \delta_i^{l+1} \cdot \theta_{i,j}^{(l)} \cdot a_j^l(1 - a_j^l)$, where $l = L - 1, L - 2, ..., 2$
- We still get, $\frac{\partial J(\theta)}{\partial \theta_{i,j}^{(l)}} = \delta_i^{l+1} \cdot a_j^l$

Step 4: Update each $\theta_{i,j}^{(l)}$ (weights)

- $\theta_{i,j}^{(l)} = \theta_{i,j}^{(l)} - \alpha \cdot \frac{\partial J(\theta)}{\partial \theta_{i,j}^{(l)}}$
- $\theta_{i,j}^{(l)} = \theta_{i,j}^{(l)} - \alpha \cdot (\delta_i^{l+1} \cdot a_j^l)$, where
  - $\theta_{i,j}^{(l)}$: The parameter (e.g., weight) connecting node (j) in layer (l) to node (i) in layer $l+1$.
  - α: The learning rate.
  - δ_i^{l+1}: The error term (delta) for node (i) in layer $l+1$.
  - a_j^l: The activation of node (j) in layer (l).

Activation Functions

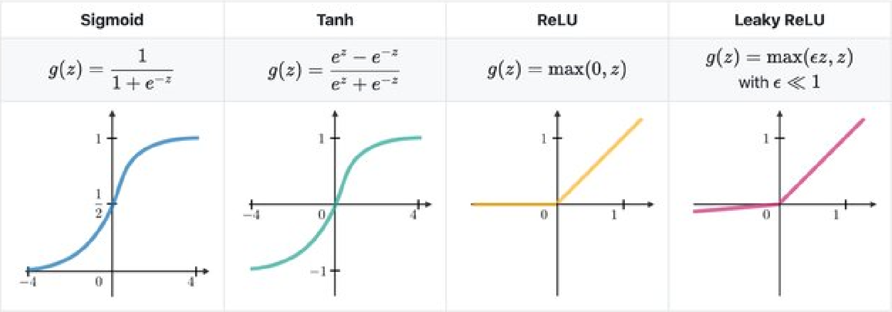


Figure: Activation function Ref:
`http://dx.doi.org/10.13140/RG.2.2.14692.60800`

Softmax Functions

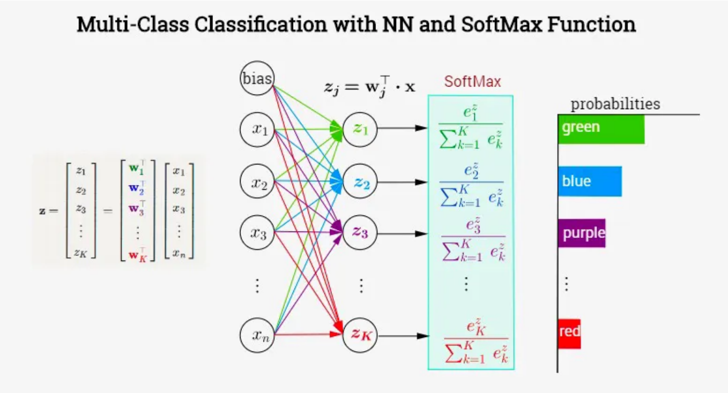
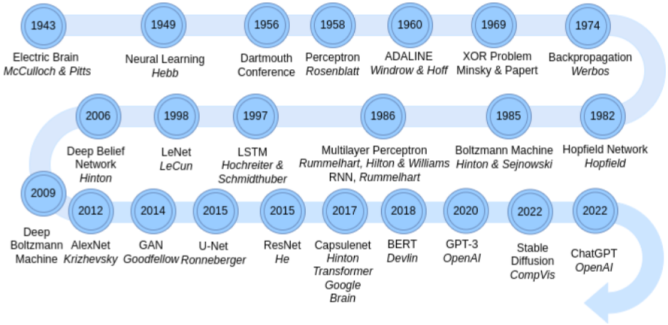


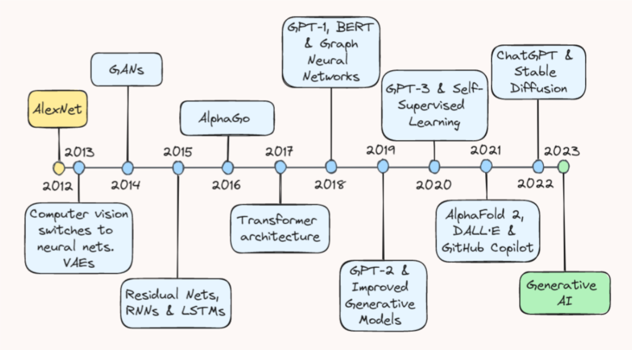
Figure: Softmax function Ref:
`https://adeveloperdiary.com/data-science/deep-learning/neural-network-with-softmax-in-python/`

History of Neural Networks (1/2)



Ref: https://pub.towardsai.net/a-brief-history-of-neural-nets-472107bc2c9c

History of Neural Networks (2/2)



Neural network zoo:
`https://www.asimovinstitute.org/neural-network-zoo/`

Reference

- Abiodun, O.I., et. al.(2018).State-of-the-art in artificial neural network applications: A survey Heliyon, 4(11).