Neural Networks - I

Thang Vu

20.11.2025

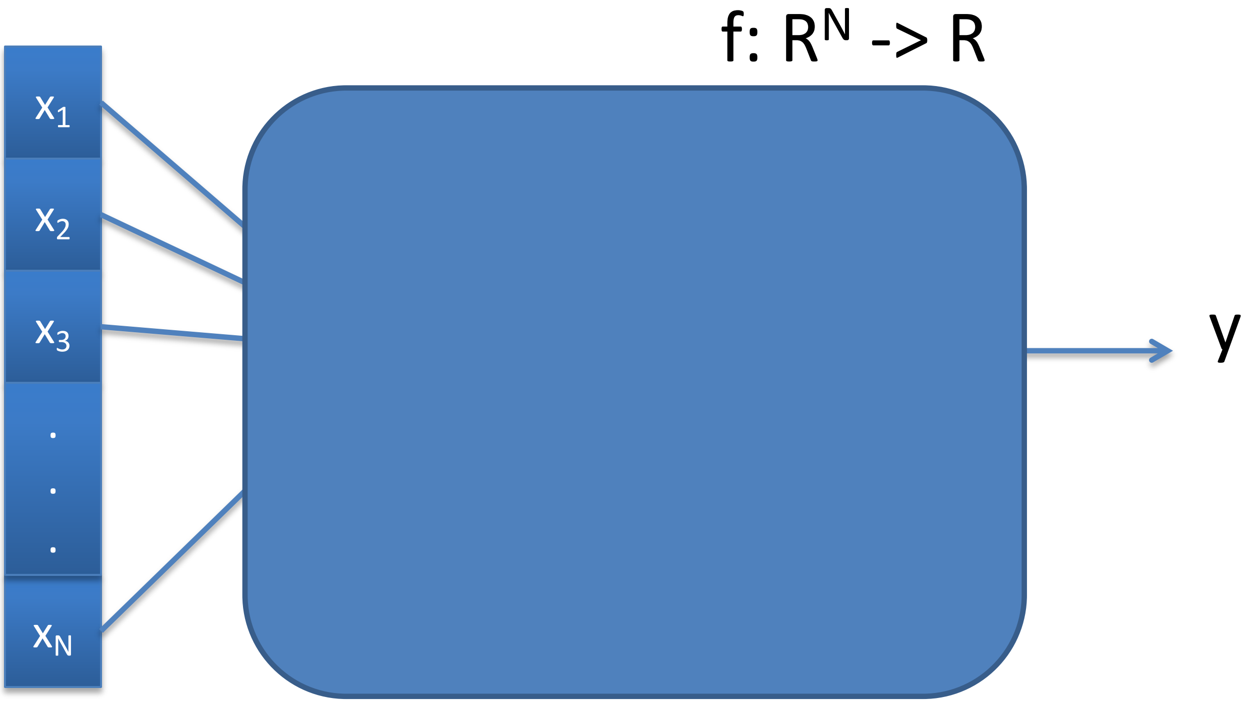
Questionnaire

- Please go to Ilias
- Open the questionnaire ‚ML Basics'
- Note that the questionnaire is anonymous, meaning we only receive the final statistics and responses, not the identities of the individuals who submitted them.

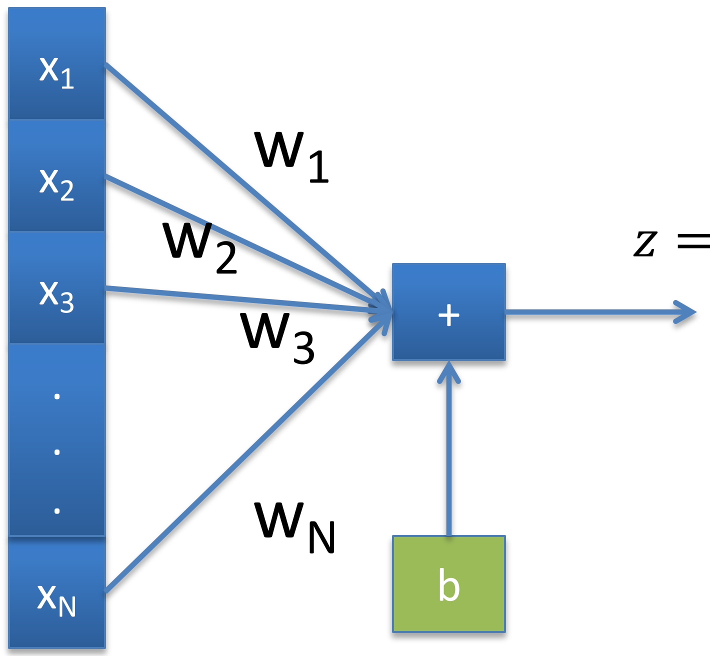
Outline

- Neuron
- Neural network
- Computing the output
- Training the network
- Backpropagation

Model: A single neuron



Model: A single neuron

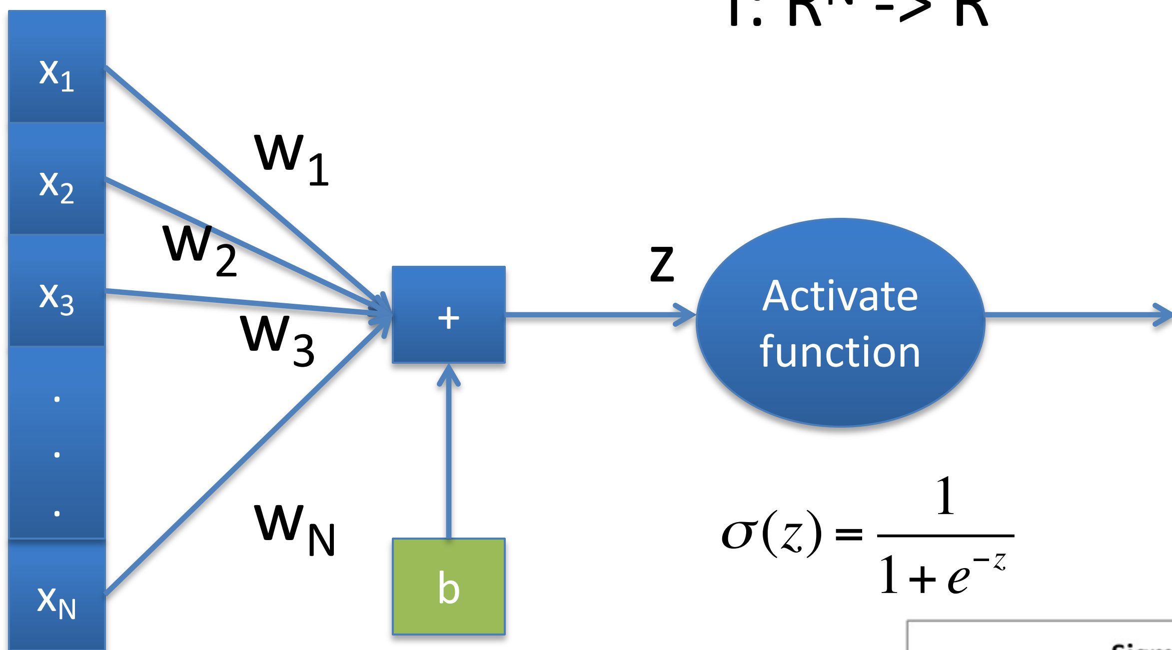


f: R^N -> R

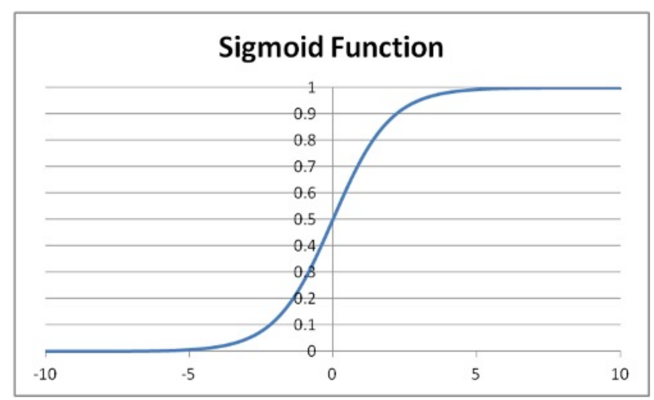
$$z = w_1x_1 + w_2x_2 + \cdots + w_nx_n + b$$

Model: A single neuron

f: $R^N \rightarrow R$

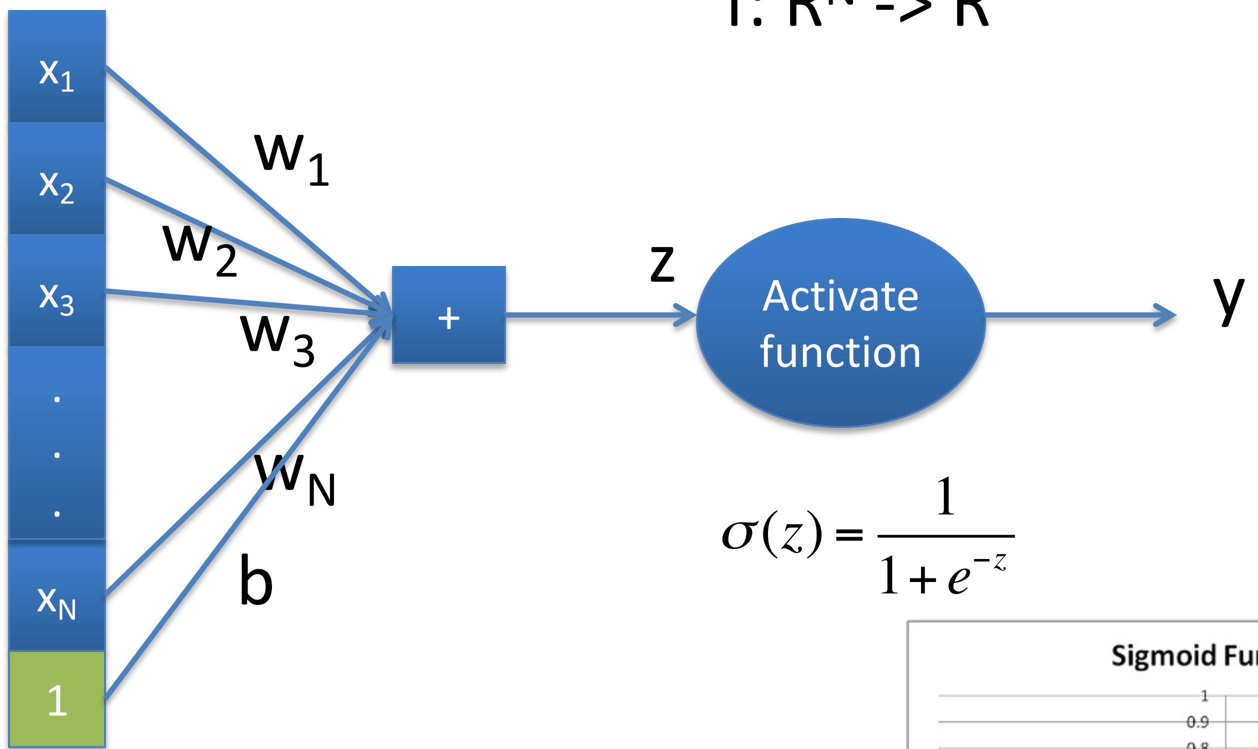


$$\sigma(z) = \frac{1}{1+e^{-z}}$$

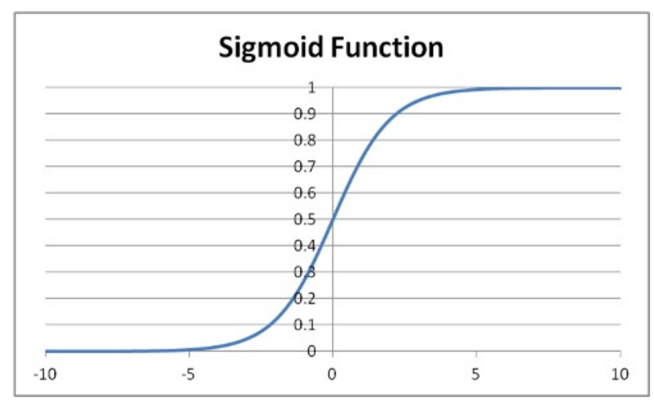


Model: A single neuron

$f: R^N \to R$

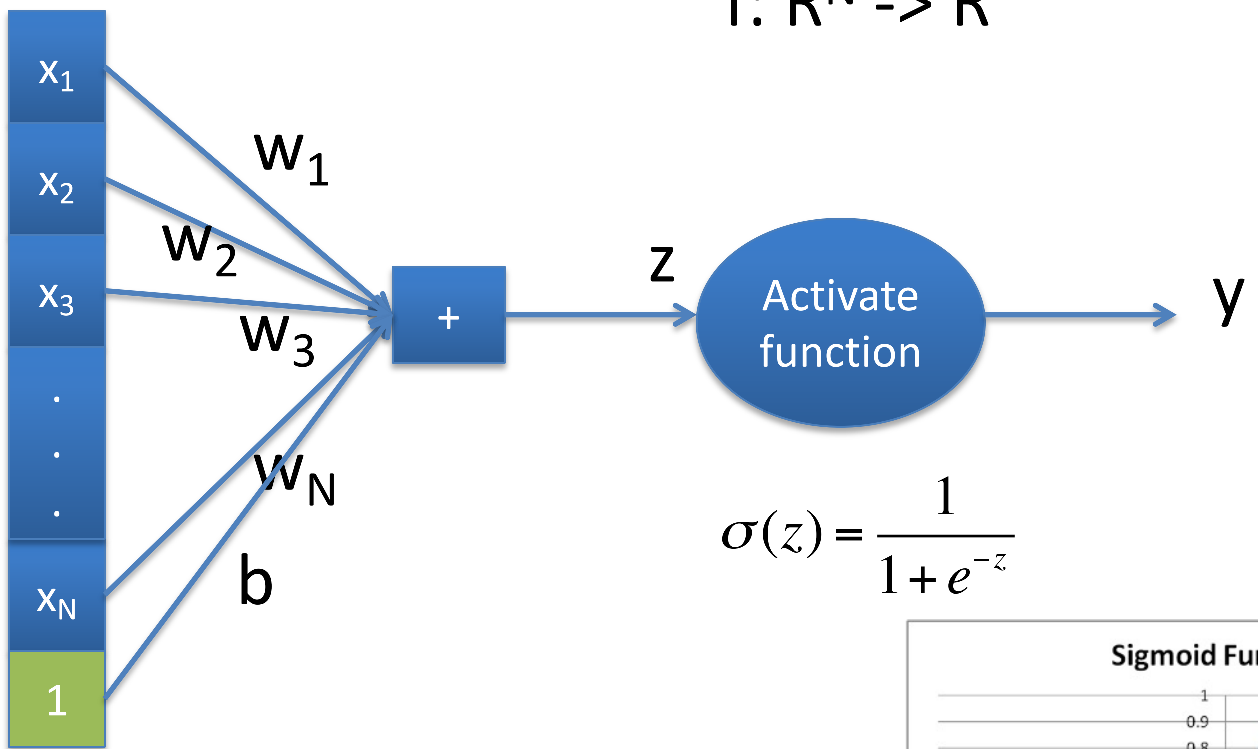


$$\sigma(z) = \frac{1}{1+e^{-z}}$$

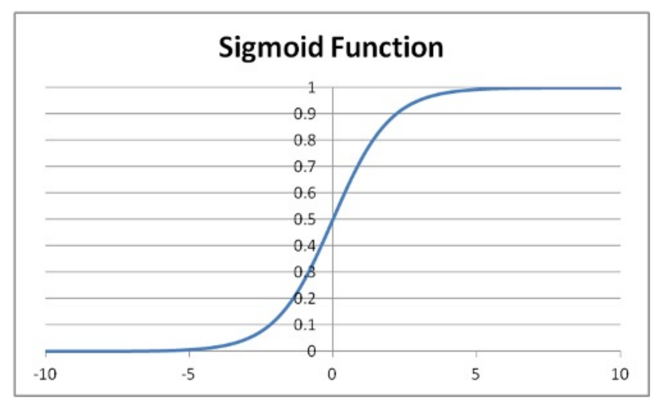


Model: A single neuron

f: R^N -> R

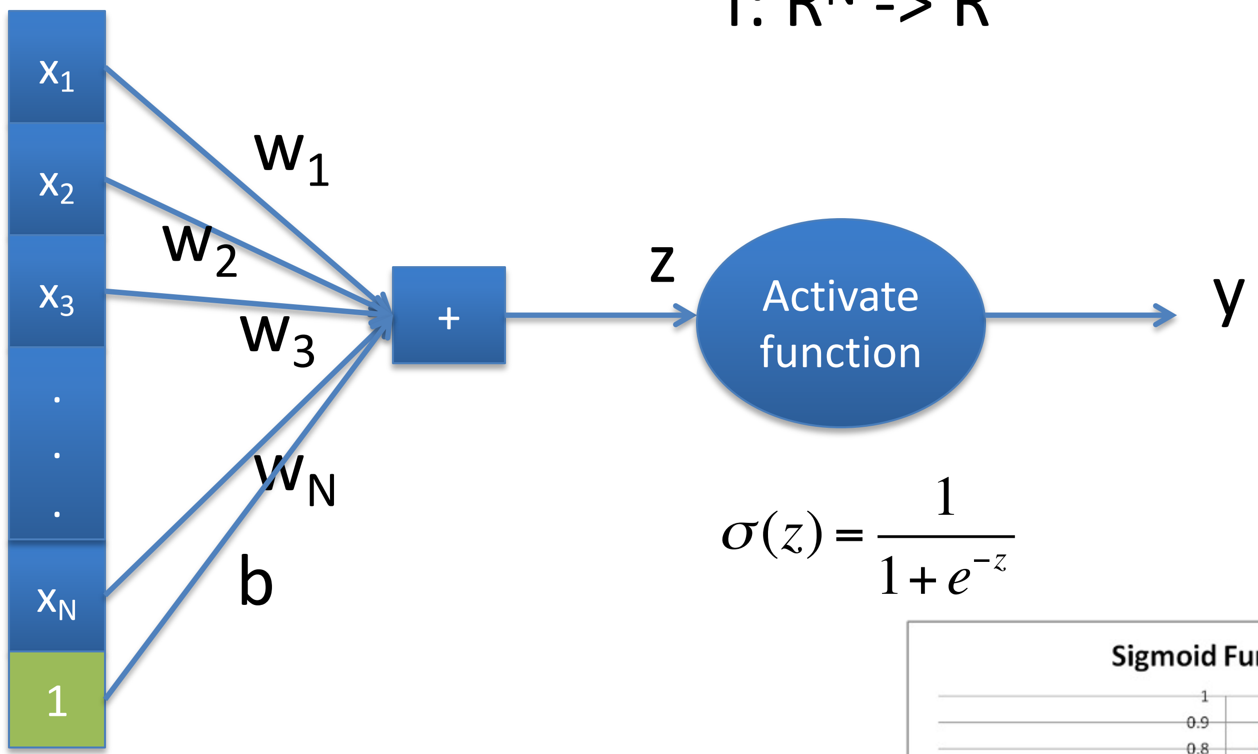


$$\sigma(z) = \frac{1}{1+e^{-z}}$$

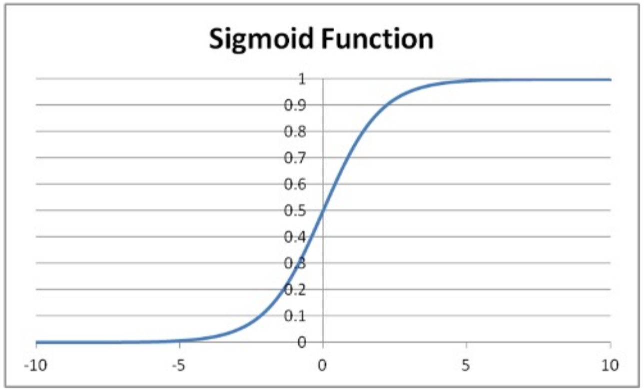


Model: A single neuron

f: R^N -> R

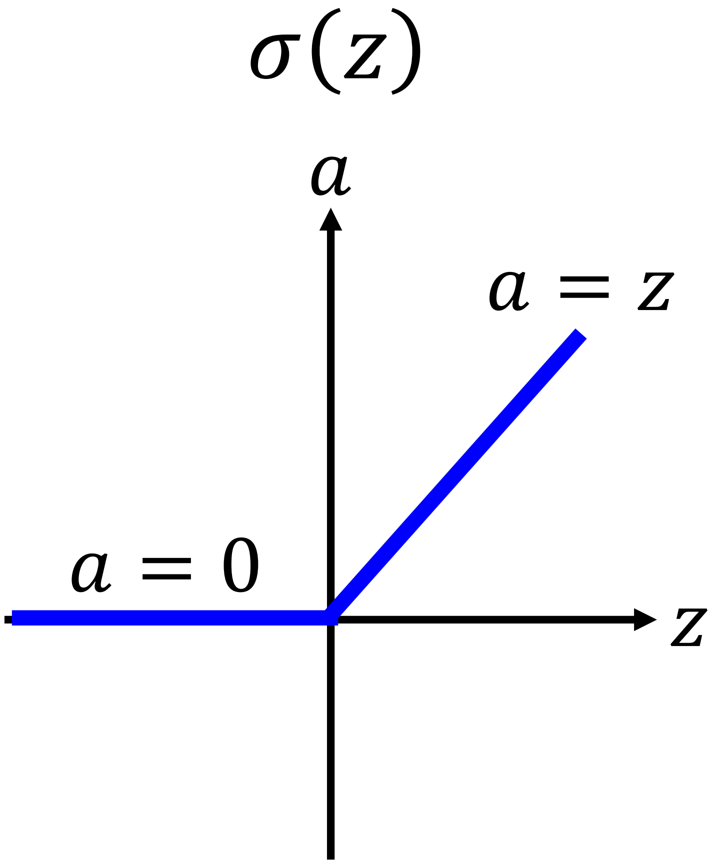


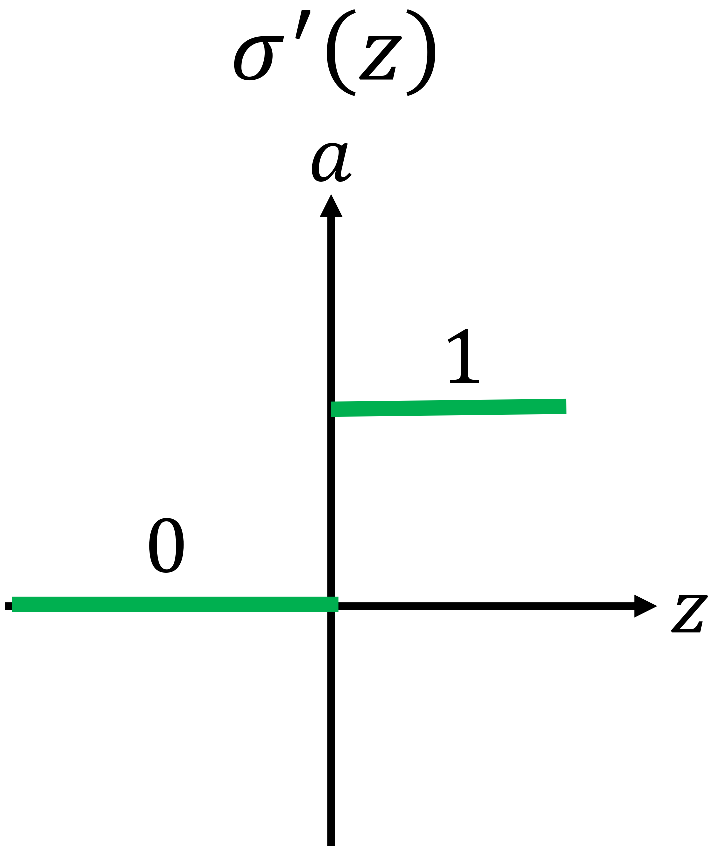
$$\sigma(z) = \frac{1}{1+e^{-z}}$$



Logistic regression model? – Yes!

Recitifier Linear Unit (ReLU)

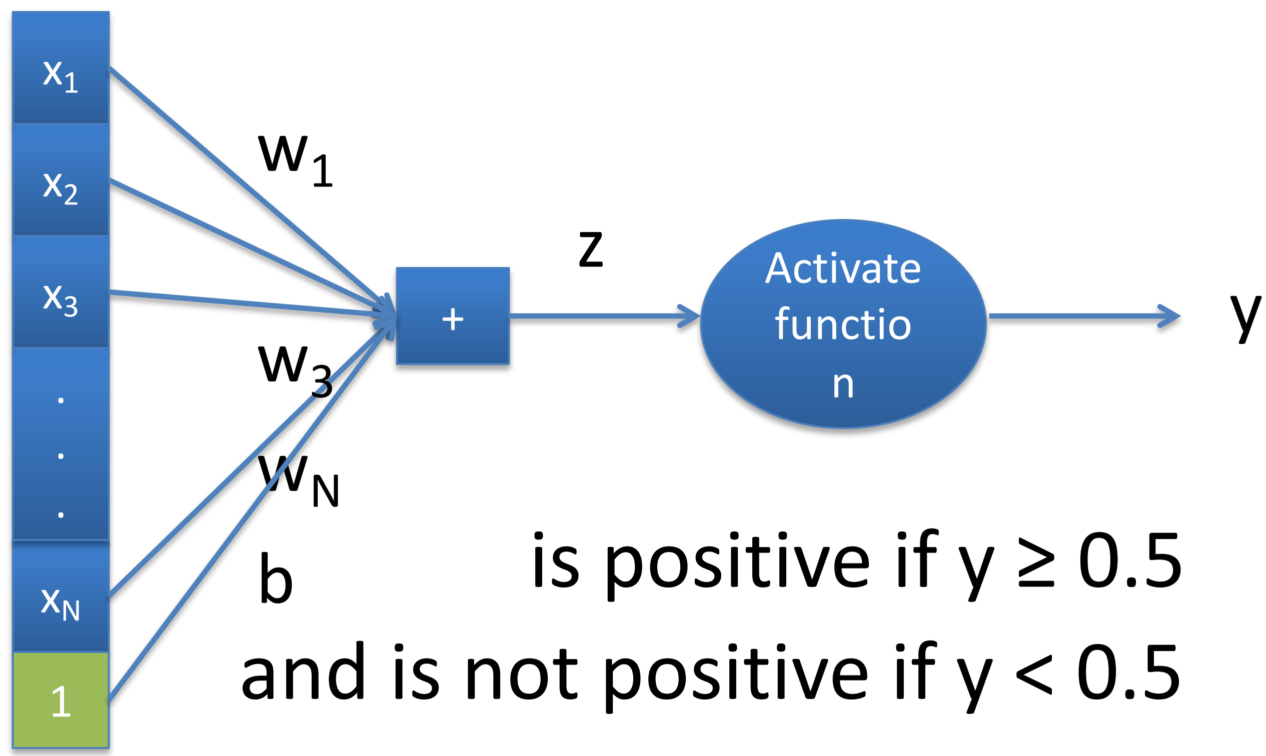




A single neuron

- A single neuron can handle binary classification task

The movie is great

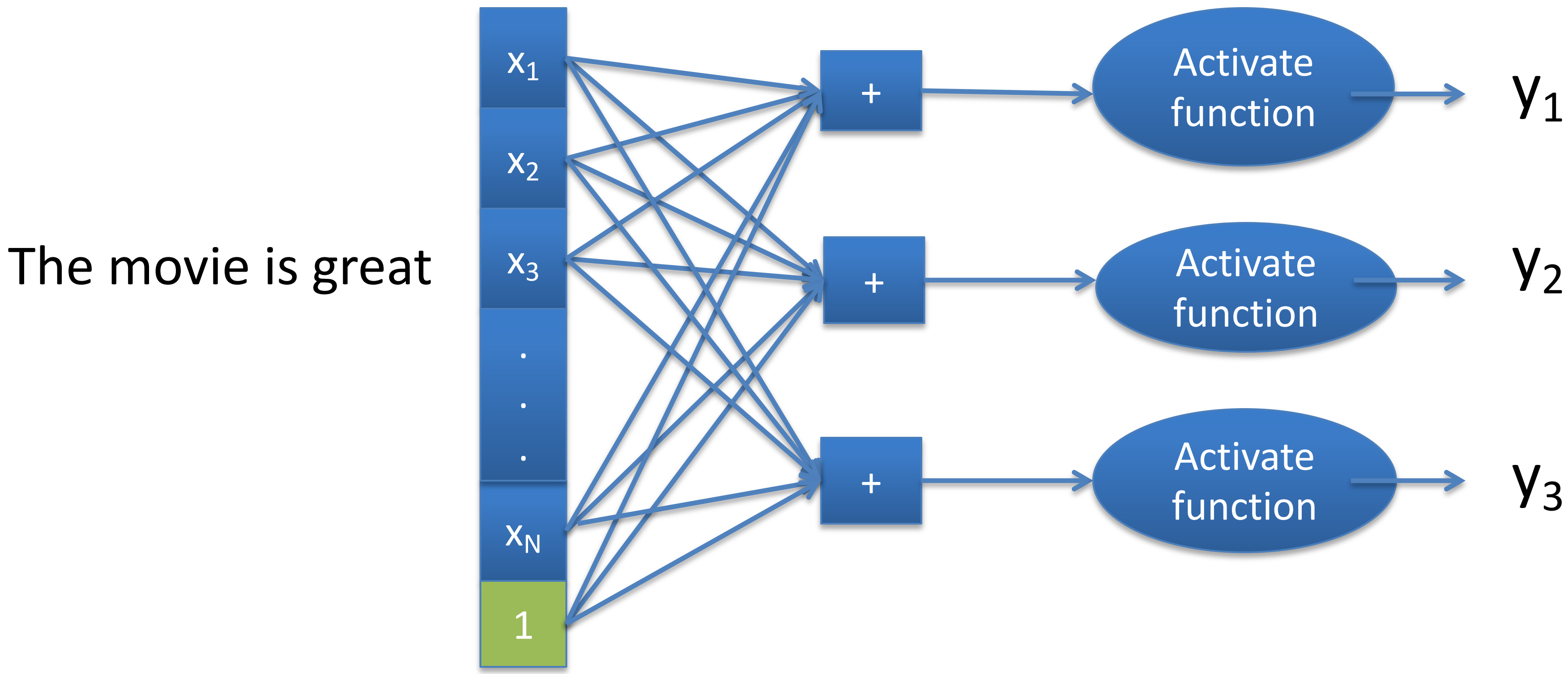


is positive if $y \geq 0.5$

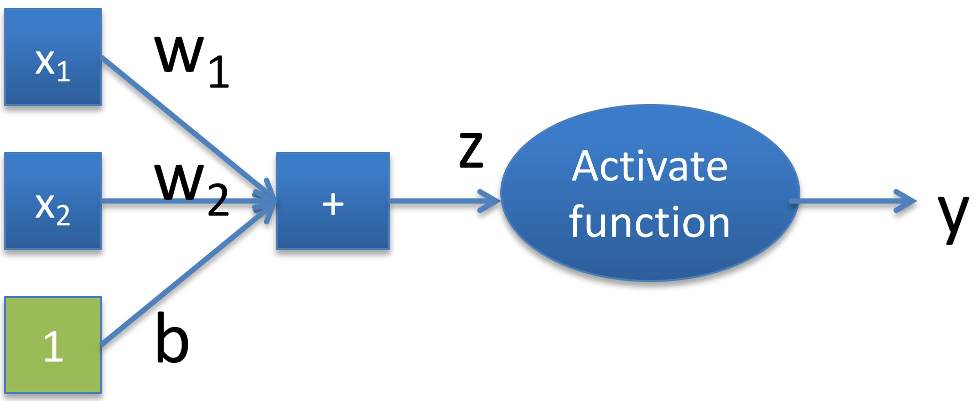
and is not positive if $y < 0.5$

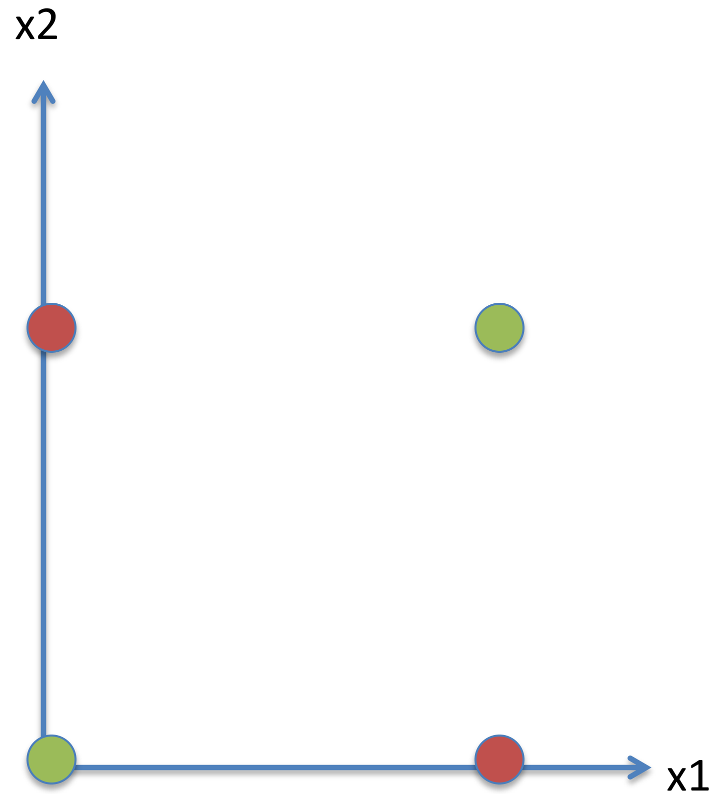
Layer of neurons

- Sentiment analysis with 3 classes
  - Positive, negative, and neutral



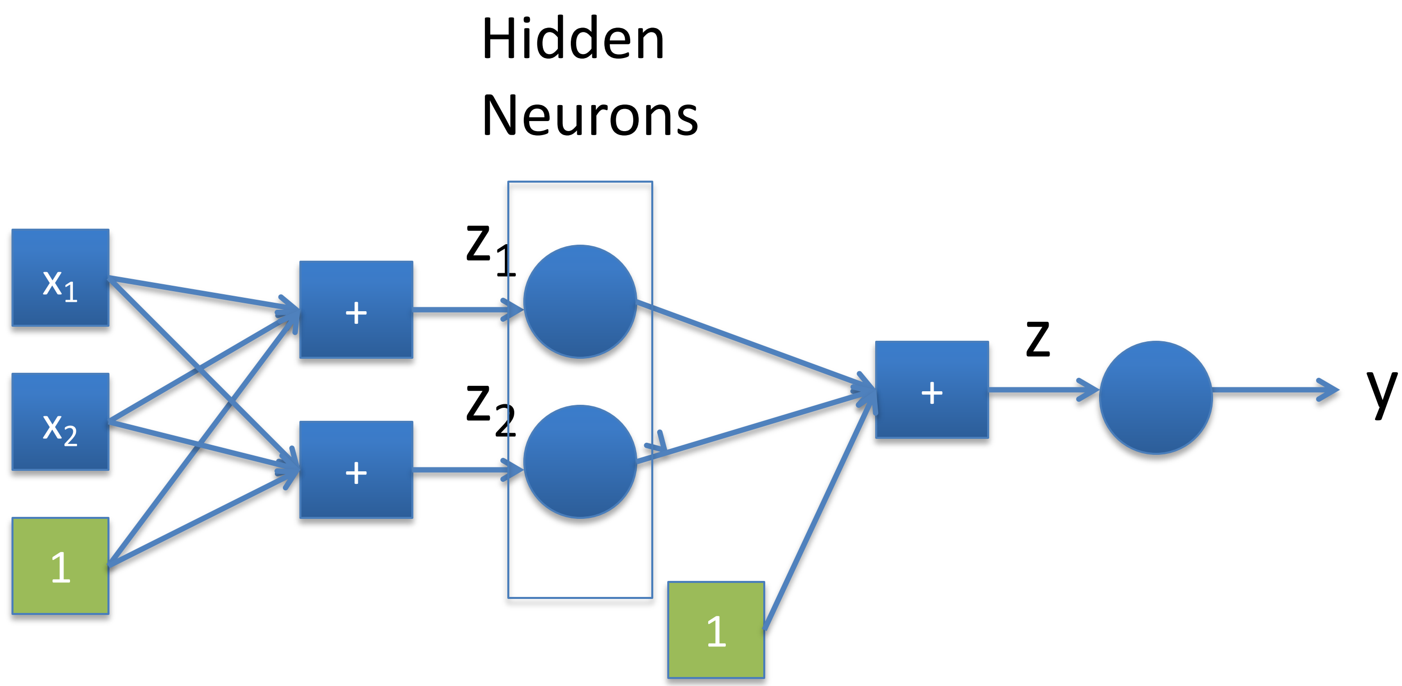
Limitation of single layer



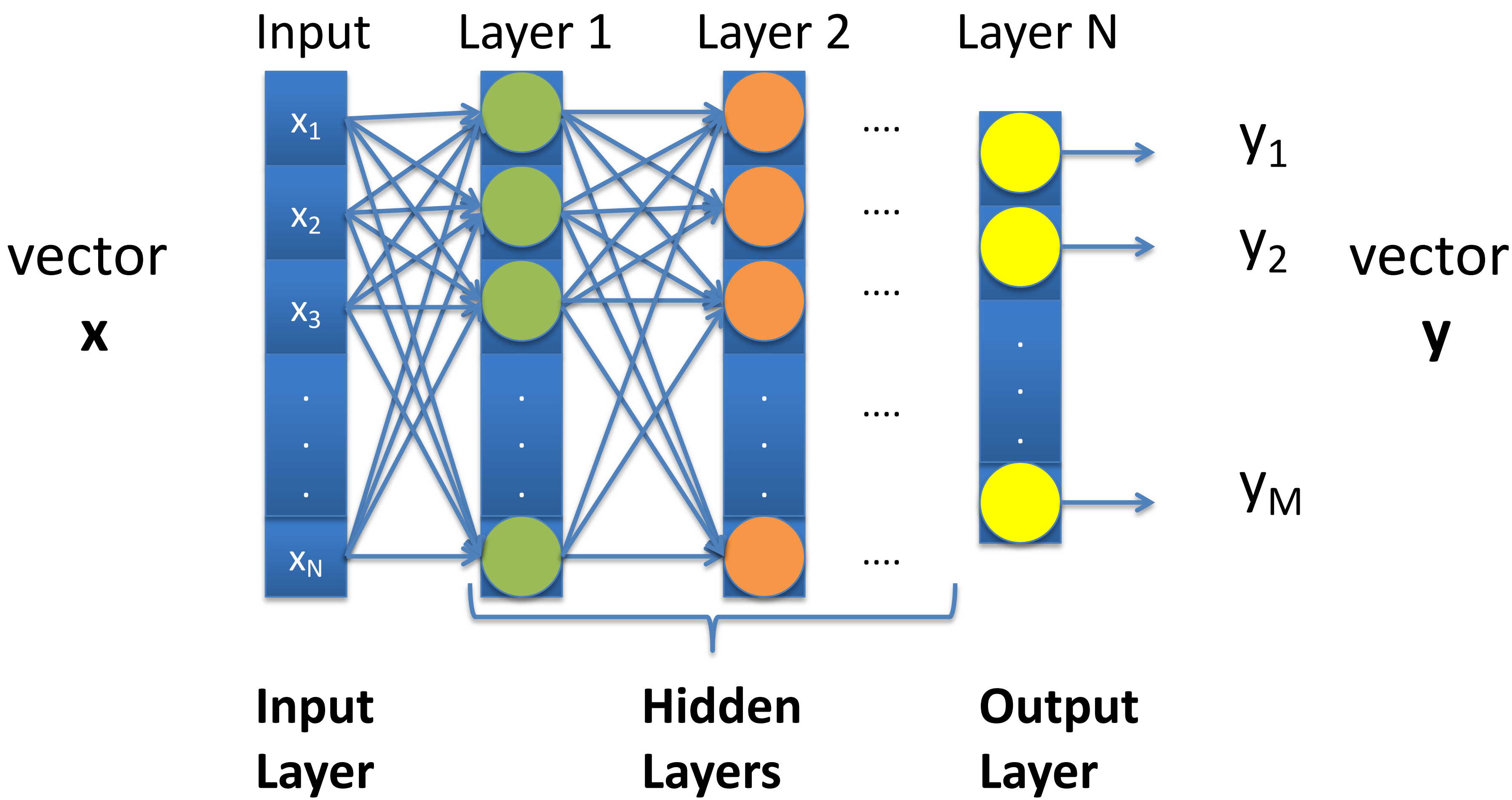


- A single layer of neurons can not handle XOR problem?

Neural network

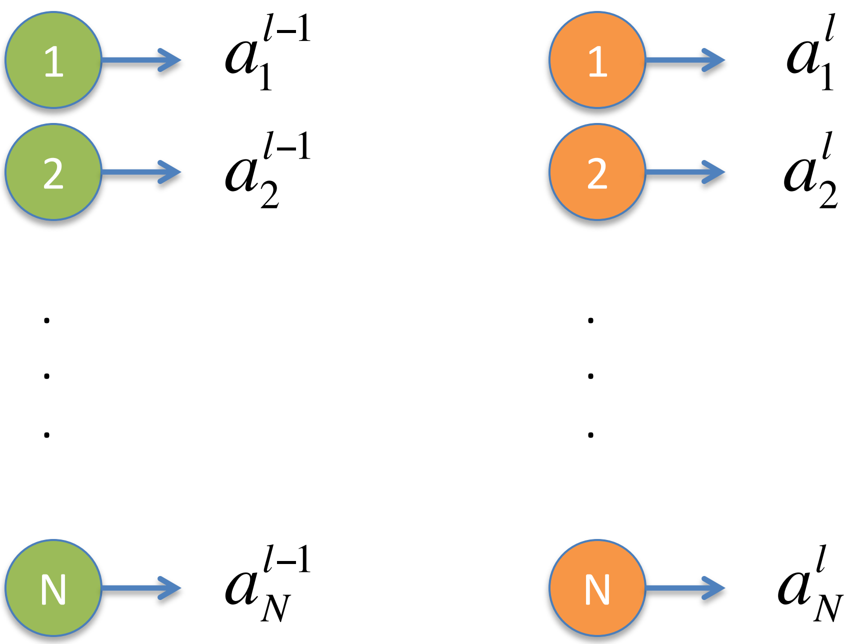


Neural network with hidden layers



Live Voting

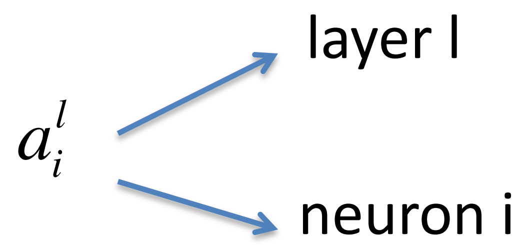
Notation



Layer l-1
N_{l-1} nodes

Layer l
N_l nodes

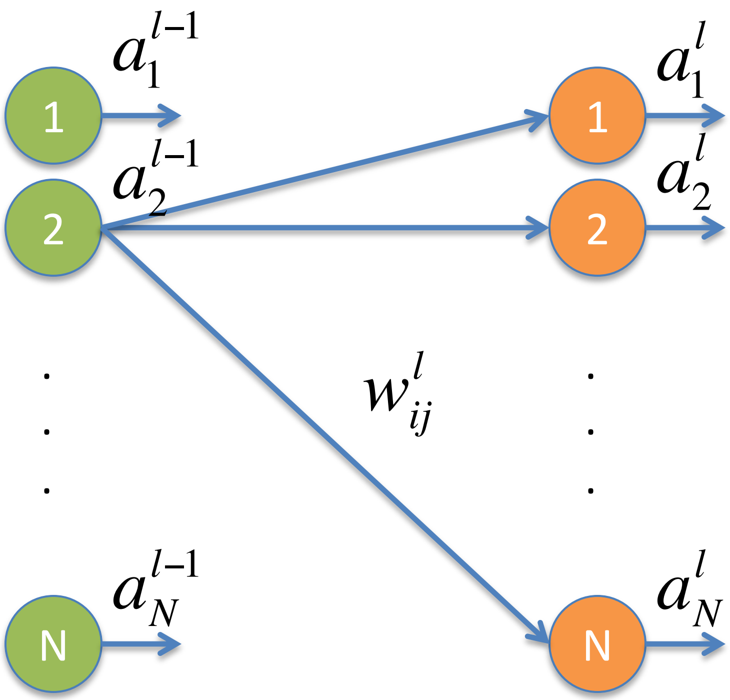
- Output of a neuron:



- Output of one layer:

a^l is a vector

Notation



Layer l-1

N_{l-1} nodes

Layer l

N_l nodes

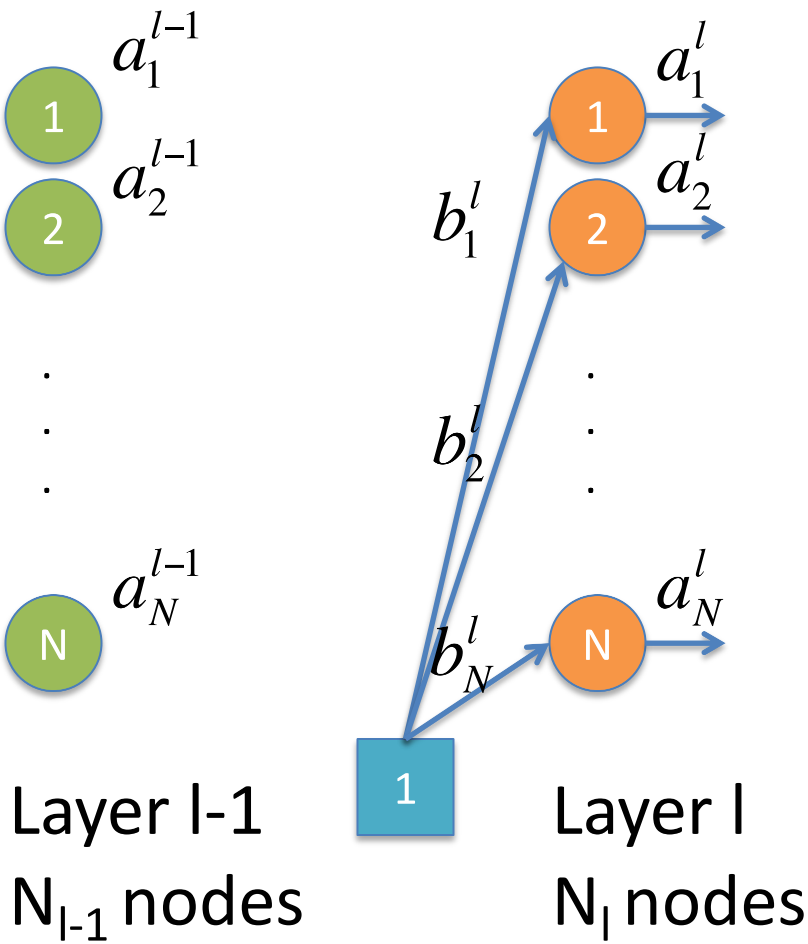
- **Weights:**

w_{ij}^l → layer *l* – *1* to layer *l*

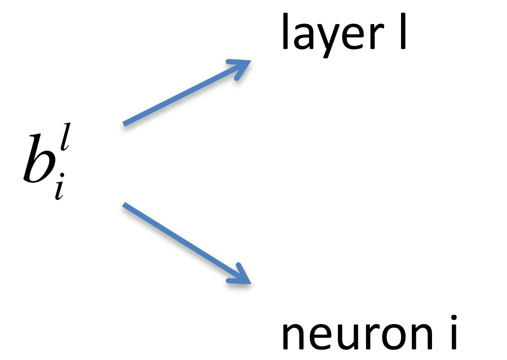
→ from neuron *j* (layer *l* -*1*) to neuron *i* (layer *l*)

$$W^l = \underbrace{\begin{bmatrix} w_{11}^l & w_{12}^l & \cdots \\ w_{21}^l & w_{22}^l & \cdots \\ \vdots & & \end{bmatrix}}_{N_{l-1}} \Big\} N_l$$

Notation



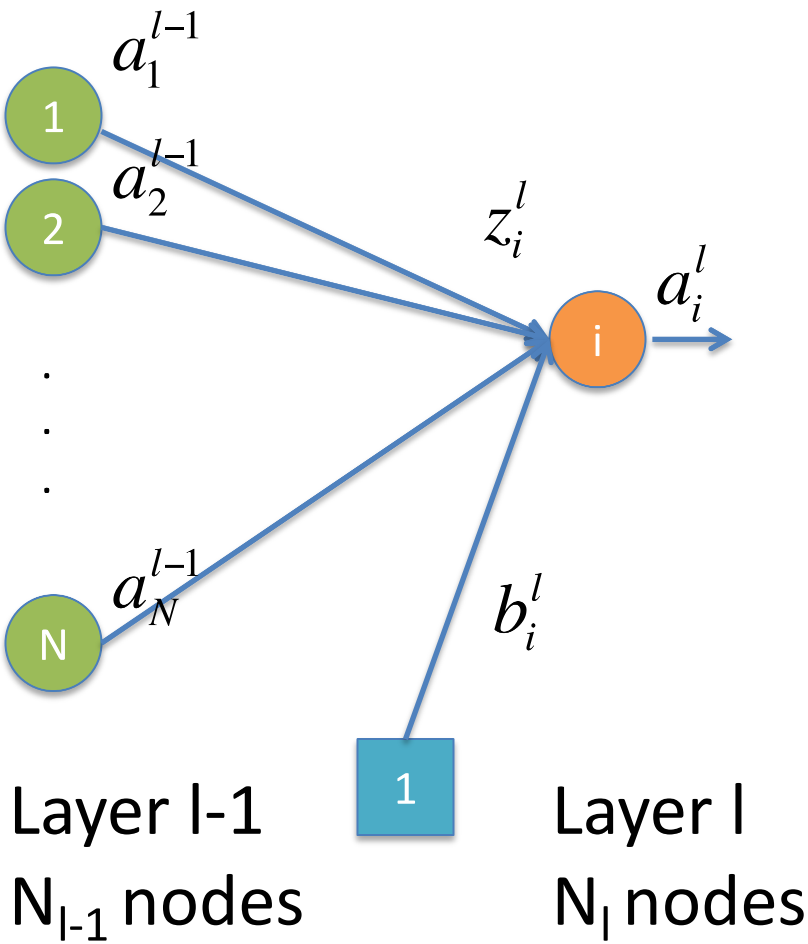
- Biases:



$$b^l = \begin{bmatrix} b_1^l \\ b_2^l \\ \vdots \end{bmatrix}$$

Bias for all the neurons in layer l

Notation



z_i^l : input of the activation function for neuron i at layer l

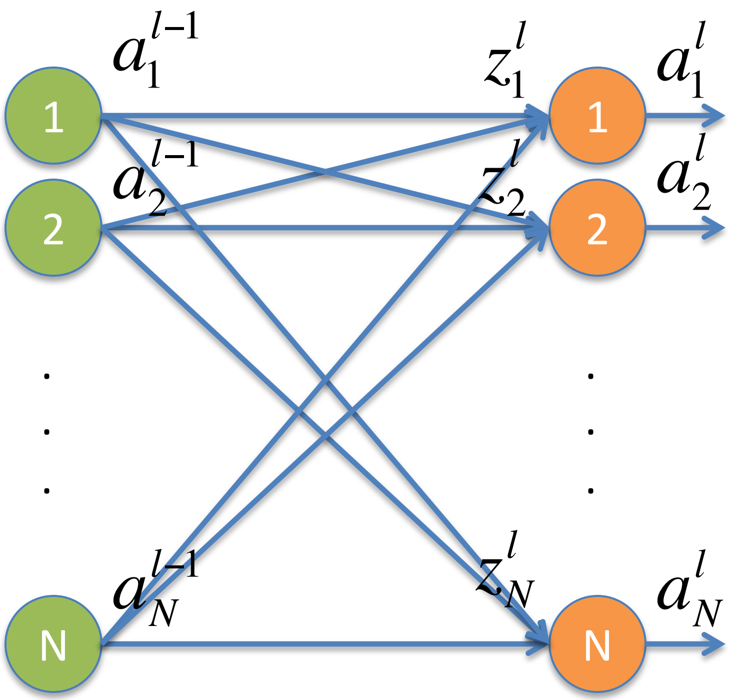
z^l : input of the activation function of all the neurons in layer l

$$z_i^l = w_{i1}^l a_1^{l-1} + w_{i2}^l a_2^{l-1} + \ldots + b_i^l$$

or in another form

$$z_i^l = \sum_{j=1}^{N_{l-1}} w_{ij}^l a_j^{l-1} + b_i^l$$

Relations between layer outputs



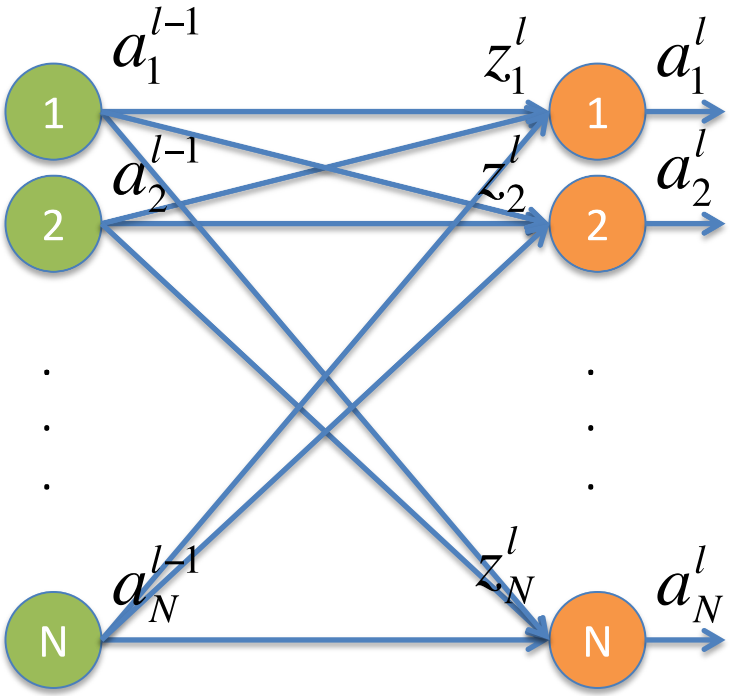
Layer l-1

N_{l-1} nodes

Layer l

N_l nodes

Relations between layer outputs



$$z_1^l = w_{11}^l a_1^{l-1} + w_{12}^l a_2^{l-1} + \ldots + b_1^l$$

$$z_2^l = w_{21}^l a_1^{l-1} + w_{22}^l a_2^{l-1} + \ldots + b_2^l$$

.....

$$z_N^l = w_{N1}^l a_1^{l-1} + w_{N2}^l a_2^{l-1} + \ldots + b_N^l$$

$$z^l = W^l a^{l-1} + b^l$$

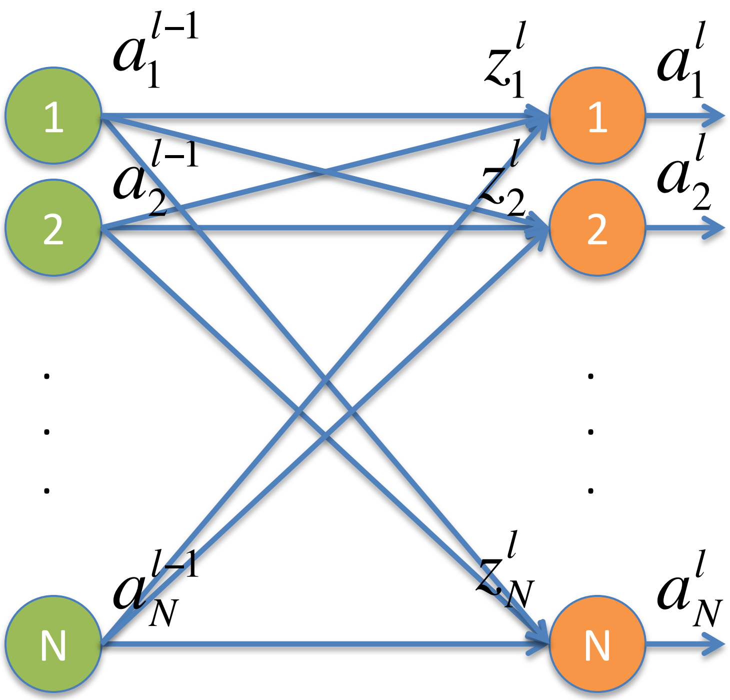
Layer l-1

N_{l-1} nodes

Layer l

N_l nodes

Relations between layer outputs



Layer l-1

N_{l-1} nodes

Layer l

N_l nodes

$$a_1^l = \sigma(z_1^l)$$

$$a_2^l = \sigma(z_2^l)$$

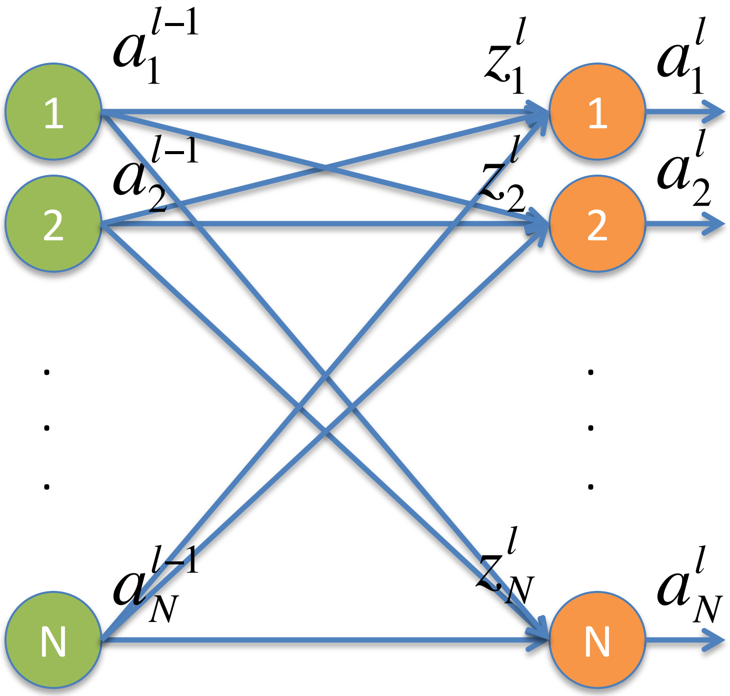
....

$$a_N^l = \sigma(z_N^l)$$

⬇

$$a^l = \sigma(z^l)$$

Relations between layer outputs



Layer l-1
N_{l-1} nodes

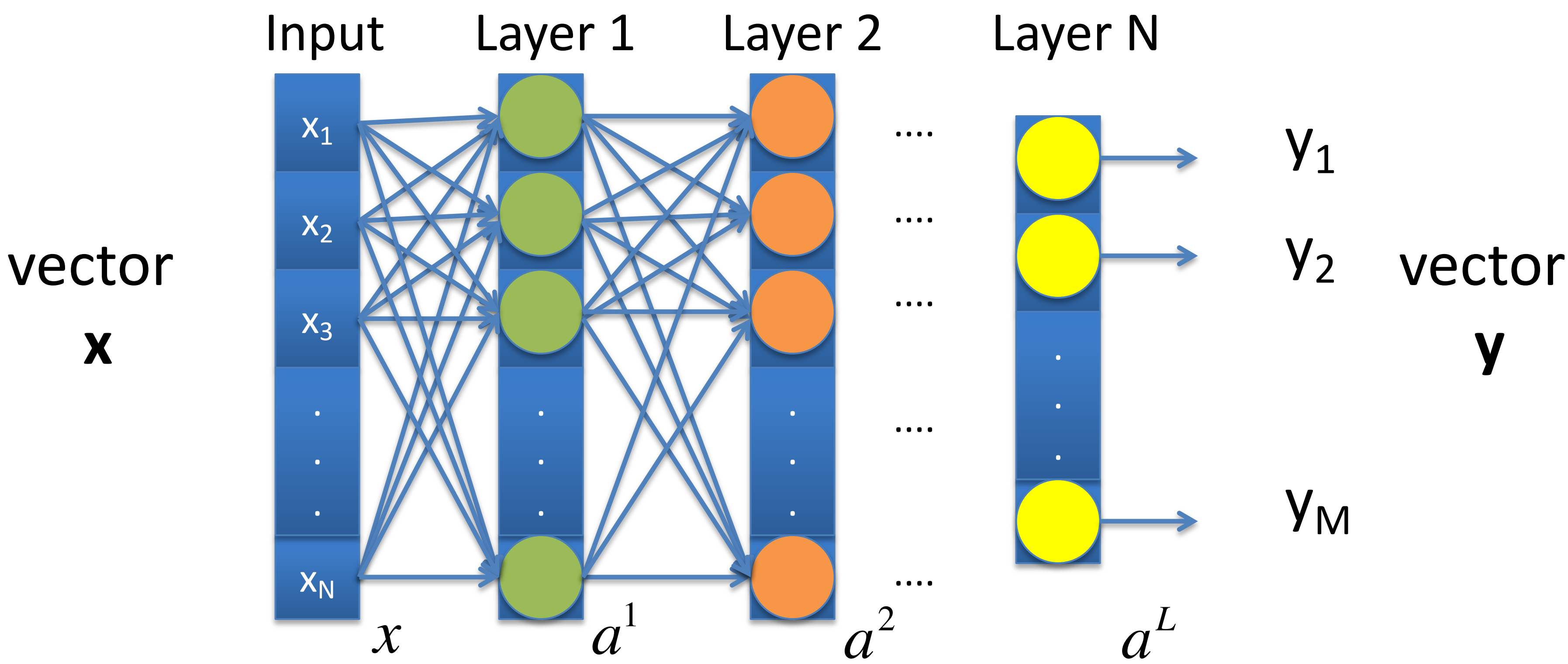
Layer l
N_l nodes

$$z^l = W^l a^{l-1} + b^l$$

$$a^l = \sigma(z^l)$$

$$a^l = \sigma(W^l a^{l-1} + b^l)$$

Computation of the final output

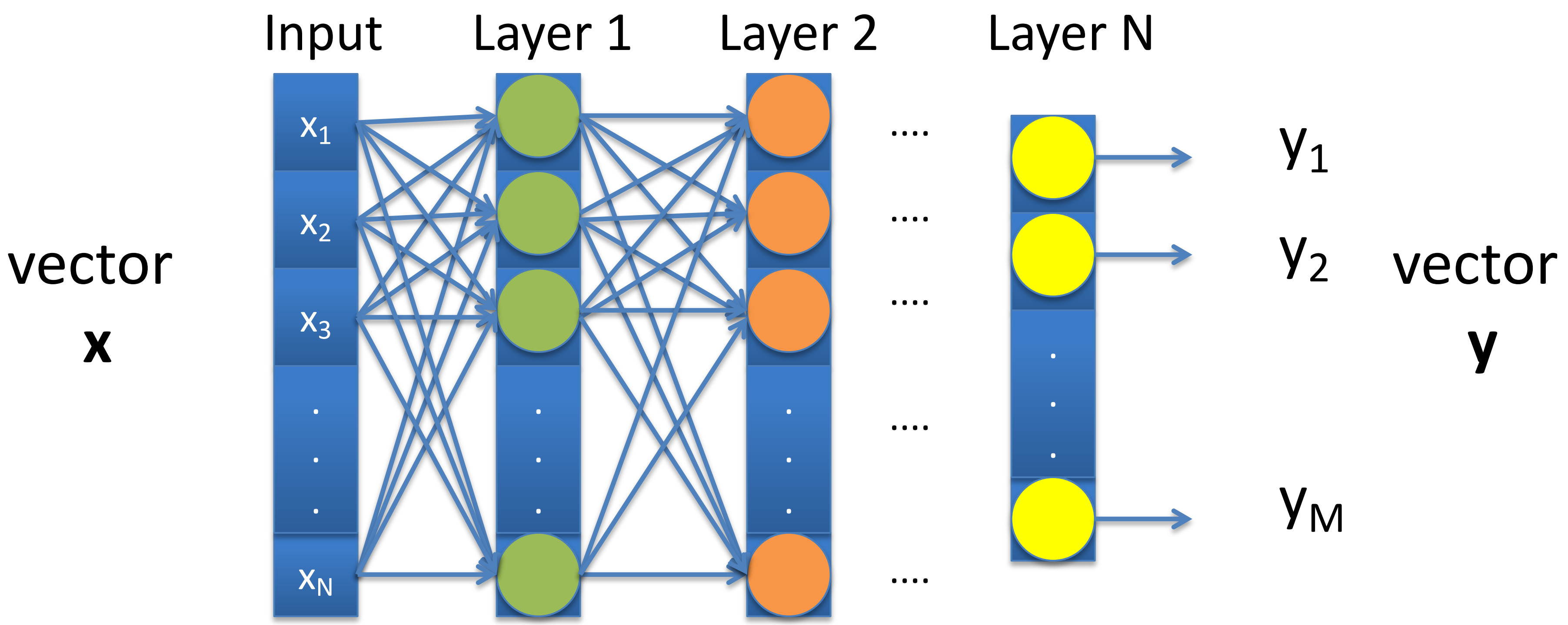


$$a^1 = \sigma(W^1 x + b^1)$$

$$a^L = \sigma(W^L a^{L-1} + b^L)$$

$$a^2 = \sigma(W^2 a^1 + b^2)$$

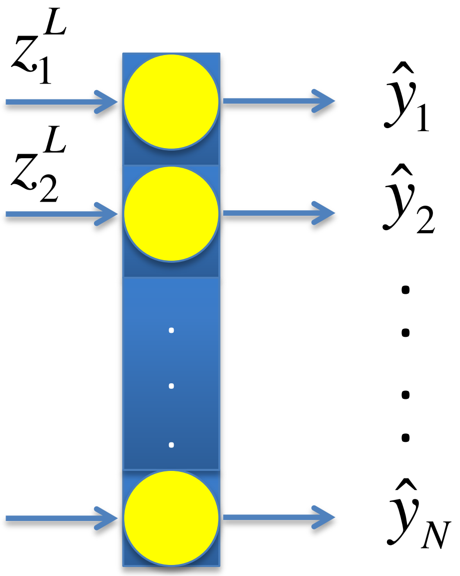
Computation of the final output



$$y = f(x) = \sigma(W^L \ldots \sigma(W^2 \sigma(W^1 x + b^1) + b^2) \ldots + b^L)$$

Softmax function

Ouput layer L



- Outputs are probabilities

$$0 < y_i < 1$$

$$\sum_i y_i = 1$$

- Sofmax function:

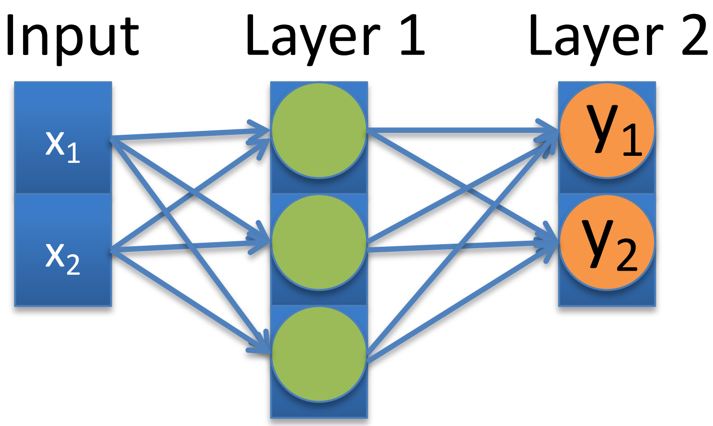
$$y_i = \frac{e^{z_i^L}}{\sum_j e^{z_j^L}}$$

$$\frac{\delta y_i}{\delta z_j} = y_j(1 - y_j) \quad i = j$$

$$= -y_i y_j \quad i \neq j$$

Exercise 1

- Compute the output **y** of the following network
- Layer 1 uses ReLU
- Layer 2 uses Softmax



$$x = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$W^1 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ -1 & 0 \end{bmatrix} \quad W^2 = \begin{bmatrix} 1 & 1 & -1 \\ 0 & 1 & 0 \end{bmatrix}$$

$$b^1 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \quad b^2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Live Voting

Thanks for listening!