Neural Networks

Miriam Wagner

IDS-I5

Chair of Process and Data Science

RWTH AACHEN UNIVERSITY

Network Layout

Exercise 1

Imagine a car renting company wants to deploy a new system for assessing worthiness of its customers. The new system is using a feed forward neural network as a supervised learning algorithm. We want to classify potential company customers as good customers or bad ones depending on the price of the rented car. Therefore, we have a training dataset describing past customers using the following attributes:

- **Age {[18..30], [30..50], [50..65],[65+]}**
- **Marital status {married, single, divorced}**
- **Gender {male, female}**
- **Income {[10K..25K], [25K..50K], [50K..65K], [65K..100K], [100K+]}.**

What is the layout of the network that should be used?

Chair of Process and Data Science

Network Layout

Exercise 2

Given the logic formula $(A \vee \neg B)\mathrm{XOR}(\neg C \vee \neg D)$.

Assume as input 0 for FALSE and 1 for TRUE.

$\longrightarrow$ **Create a neural network with one hidden layer that implements the truth value of the formula. Draw your network and show all weights. As activation function use the step function with threshold 1.**

Recap
Two layer neural network

N output neurons:
$1 \leq k \leq N$

Number of neurons in the hidden layer

Number of inputs

Weight's layer

$$y_k(x,w) = f\left(\sum_{j=0}^{M} w_{jk}^{(2)} h\left(\sum_{i=0}^{D} w_{ij}^{(1)} x_i\right)\right)$$

- f and h are activation functions:
  - they can be different functions

    $\rightarrow$ for simplicity we assume there is one h function for the hidden layer
- w_{jk} shows the weight of the edge from the $jth-$neuron of the hidden layer to the $kth-$neuron of the output layer
- w_{ij} shows the weight of the edge from the $ith-$element of the input to the $jth-$neuron of the hidden layer

Chair of Process and Data Science

Feedforward

Exercise 1

Consider three inputs for a neuron with the following weights and activation function:

$$f(x) = \begin{cases} \mathbf{0}, & x < 0 \\ \mathbf{1}, & x \geq 0 \end{cases}$$

Weight	Value
w_0	0
w_1	2
w_2	-4
w_3	1

Given the input patterns p_1 to p_4 below, calculate the output of the neuron.

	p_1	p_2	p_3	p_4
x_1	1	0	1	1
x_2	0	1	0	1
x_3	0	1	1	1

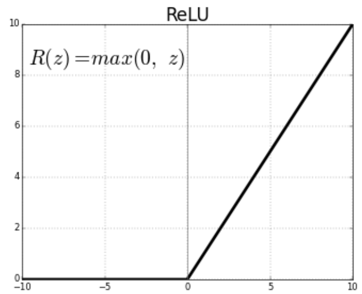
Chair of Process and Data Science

Feedforward

Exercise 2

Assume a two-layer network with the ReLU function as activation function in the hidden layer and no activation function in the output layer.

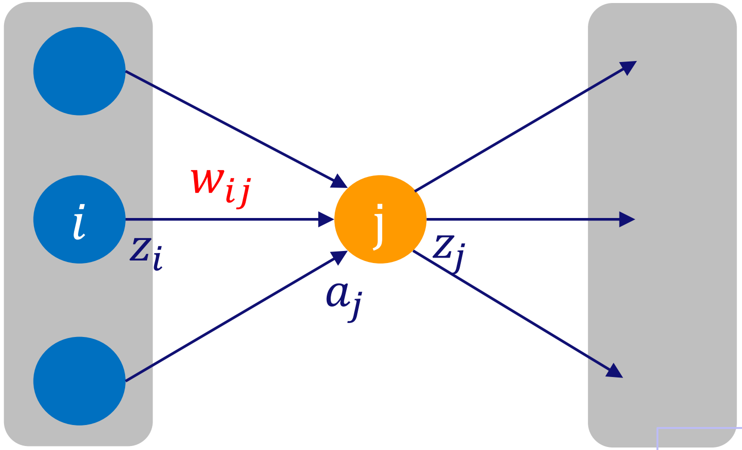
1. **Write down the equation of the output of the *jth* neuron in the hidden layer.**
2. **Write down the equation of the output of the *mth* neuron in the output layer.**



Chair of Process and Data Science

Recap

Context of a single neuron



$$a_j = \sum_i w_{ij} z_i \qquad z_j = \sigma(a_j) = \frac{1}{1 + e^{-a_j}}$$

$$Error = \frac{1}{2} \sum_k (z_k - t_k)^2$$

$$= -\frac{\partial Error}{\partial a_j} \frac{\partial a_j}{\partial w_{ij}} = -\frac{\partial Error}{\partial z_j} \frac{\partial z_j}{\partial a_j} \frac{\partial a_j}{\partial w_{ij}}$$

$E_{ij} = -\frac{\partial Error}{\partial w_{ij}}$ **shows the direction of the desired change for** w_{ij}

learning rate

$$w_{ij}^{new} = w_{ij}^{old} + \Delta w_{ij} \qquad \Delta w_{ij} = -\frac{\partial Error}{\partial w_{ij}} l = l \, E_{ij}$$

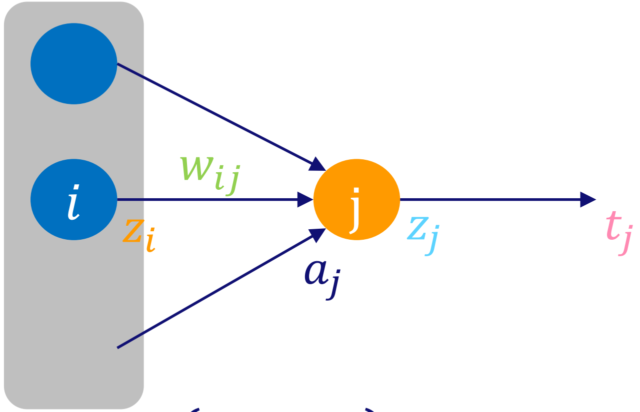
Chair of Process and Data Science

Recap

Summary of notations

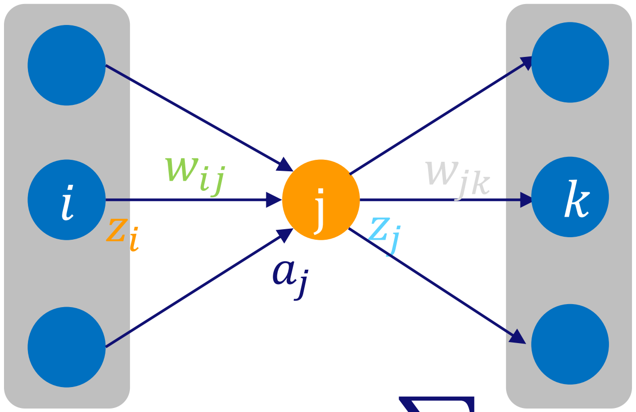
$$a_j = \sum_i w_{ij} z_i \qquad z_j = \sigma(a_j) = \frac{1}{1 + e^{-a_j}} \qquad Error = \frac{1}{2} \sum_k (z_k - t_k)^2 \qquad E_j = -\frac{\partial Error}{\partial z_j} z_j(1 - z_j)$$

$$w_{ij}^{new} = w_{ij}^{old} + \Delta \mathrm{w_{ij}} \qquad \Delta \mathrm{w_{ij}} = l\, E_{ij} = l\, E_j z_i$$



$$E_j = z_j(1 - z_j)(t_j - z_j)$$

Output layer



$$E_j = z_j(1 - z_j) \sum_k w_{jk} E_k$$

Error from layer before going backwards

Hidden layer

Chair of Process and Data Science

Recap

Weight updating

$$w_{ij}^{new} = w_{ij}^{old} + \boxed{\Delta \mathrm{w_{ij}}}$$

updated weight of connection from neuron i to neuron j based on some instance

$$\boxed{\Delta \mathrm{w_{ij}}} = \boxed{l\, z_i z_j (1 - z_j)} (t_j - z_j)$$

<u>Case 1</u>: neuron j is in the output layer

$$\boxed{\Delta \mathrm{w_{ij}}} = \boxed{l\, z_i z_j (1 - z_j)} \sum_k w_{jk} E_k$$

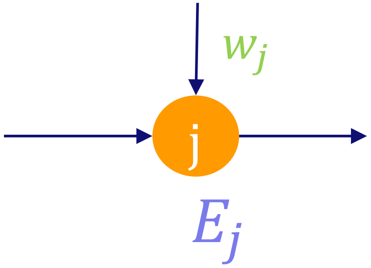
<u>Case 2</u>: neuron j is in a hidden layer

l **is a scaling parameter (learning rate)**

Clarification

Weight updating - Bias

$$w_{ij}^{new} = w_{ij}^{old} + \Delta \mathrm{w}_{\mathrm{ij}}$$



$$\Delta \mathrm{w}_{\mathrm{j}} = l \; E_j$$

Backpropagation

Exercise 1

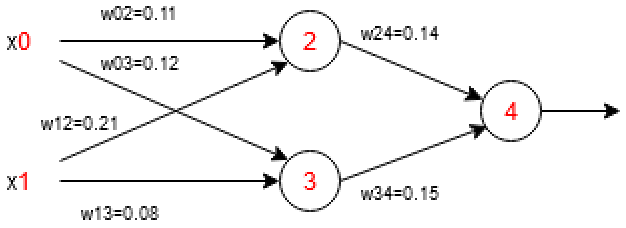
Arcs with weight 0 are not displayed.

Consider the neural network below. Assume a learning rate of 0.05 and as training data

X0	X1	Output
2	3	1

- **Consider as activation function the identity function, but**
- **consider the derivative of the activation function in the direction of a_j to be 1.**

Calculate the updated weights for one round of backpropagation.



Chair of Process and Data Science

Backpropagation

Exercise 2

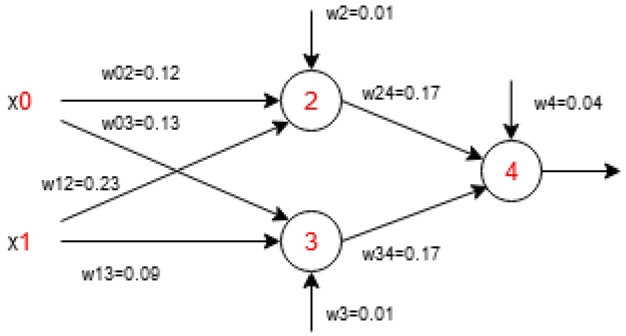
Arcs with weight 0 are not displayed.

Consider the neural network below. Assume a learning rate of 0.1 and as training data

X0	X1	Output
1	5	0

- **Consider as activation function the sigmoid function.**

Calculate the updated weights for one round of backpropagation.



Chair of Process and Data Science

Backpropagation

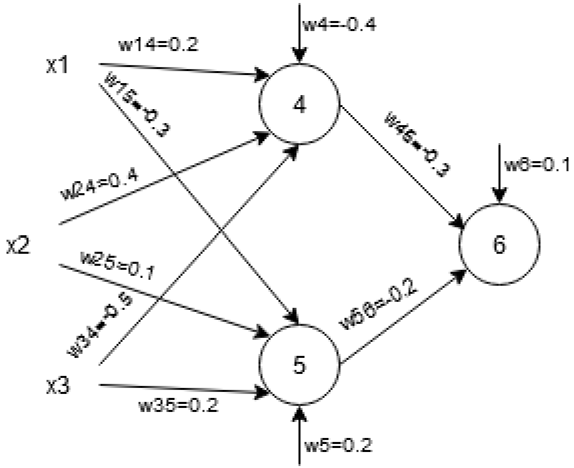
Exercise 3 (Example from the lecture)

Consider the neural network below. Assume a learning rate of 0.9 and as training data

X0	X1	X2	Output
1	0	1	1

- **Consider as activation function the sigmoid function.**

Calculate the updated weights for one round of backpropagation.



Chair of Process and Data Science

Recap

Naïve Bayes' Classifier

Naive Bayes' Classifier

$$\mathbb{M}(\mathbf{q}) = \underset{l \in levels(t)}{\operatorname{argmax}} \left(\prod_{i=1}^{m} P(\mathbf{q}[i] \mid t = l) \right) \times P(t = l)$$

- t = target feature has a specific value
- q = descriptive features have specific values
- Probabilities can be estimated in a trivial manner (count fractions of rows).

Assume independence to avoid overfitting!

Chair of Process and Data Science

Naïve Bayes classifier

Exercise 1

Age	Weight	Size	Gender [Chromosomes]
Young	Light	Small	y
Young	Average	Average	y
Young	Light	Small	x
Young	Average	Average	x
Young	Average	Average	y
Young	Heavy	Tall	y
Young	Light	Small	x
Young	Average	Average	x
Middle	Average	Average	y
Middle	Heavy	Tall	y
Middle	Average	Average	x
Middle	Heavy	Tall	x
Old	Heavy	Tall	y

Consider ‘Gender’ as the target feature and the others as descriptive features.
Give for the following input [‘Old’, ‘Average’, ‘Tall’] the gender, using a naïve Bayes classifier.

Chair of Process and Data Science