GÖTEBORGS UNIVERSITET

STUDENT

# 0002-RRF

TENTAMEN

# TIG122 Tentamen

| | |
|---|---|
| Kurskod | -- |
| Bedömningsform | DO |
| Starttid | 21.08.2024 08:30 |
| Sluttid | 21.08.2024 11:30 |
| Bedömningsfrist | -- |
| PDF skapad | 20.01.2025 15:42 |
| Skapad av | Catarina Elg |

# i General information

Welcome to TIG122_VT24 -- Artificial Intelligence -- 3rd examination session:

- Standard duration of the examination: **3 hours.**
- Number of questions: **18.**
- Number of available points: **20.** Each correctly answered question earns you one point, except for two questions, revealingly titled "Convolutional and Pooling layer" and "Vanishing gradients", which may earn you two points each.
- You will have no access to external props: no books, slides, internet, calculators, and phones.
- Question types:
  - **Multiple choice (one correct answer): 5 questions**
  - **Multiple choice (multiple possible correct answers): 2 question** --> ATTENTION: selecting the wrong answer(s) zeroes the gain from that question!
  - **Free text: 5 questions** --> Show your knowledge/understanding of the issue at hand in a relatively short form.
  - **Basic calculations: 6 questions** --> No complex calculations! In case you have to report your calculations, consider utilizing the character * to represent the multiplication operators (i.e. both for the basic multiplication and the "dot product" between matrices).

**Grading of the digital examination**

- Points ≥ 14 --> VG
- 8 ≤ Points <14 --> G
- points < 8 --> U

Warm wishes!

# 1 Turing test

Discuss the structure of the Turing test.

Is that a good way to tell if a computer program is "intelligent"? Why/why not? (**1p**)

**Fill in your answer here**

It is not, since we don't fully understand what intelligence is.

Ord: 11

Besvarad.

## 2 Decision trees

Which statement is correct? (**1p**)
**Select one alternative:**

- The output of hierarchical clustering depends on the starting point. ✗
- Bootstrapping starts by removing repeated data points.
- Random forests combine bootstrapping and limited tree size. ✓

Fel. 0 av 1 poäng.

## 3 Regression trees

Which statement is correct? (**1p**)
**Select one alternative:**

- When constructing regression trees, the goal is to minimize the Gini impurity.
- When constructing regression trees, the goal is to minimize the number of Gini dimensions.
- When constructing regression trees, the goal is to minimize the sum of the squared residuals. ✓

Rätt. 1 av 1 poäng.

## 4 ReLU activation

Consider a unit with a ReLU activation function, input x = (2,3), weights w = (−1, 2), and threshold θ = 2.
Calculate its activation a. (report your calculations) (**1p**)
**Fill in your answer here**

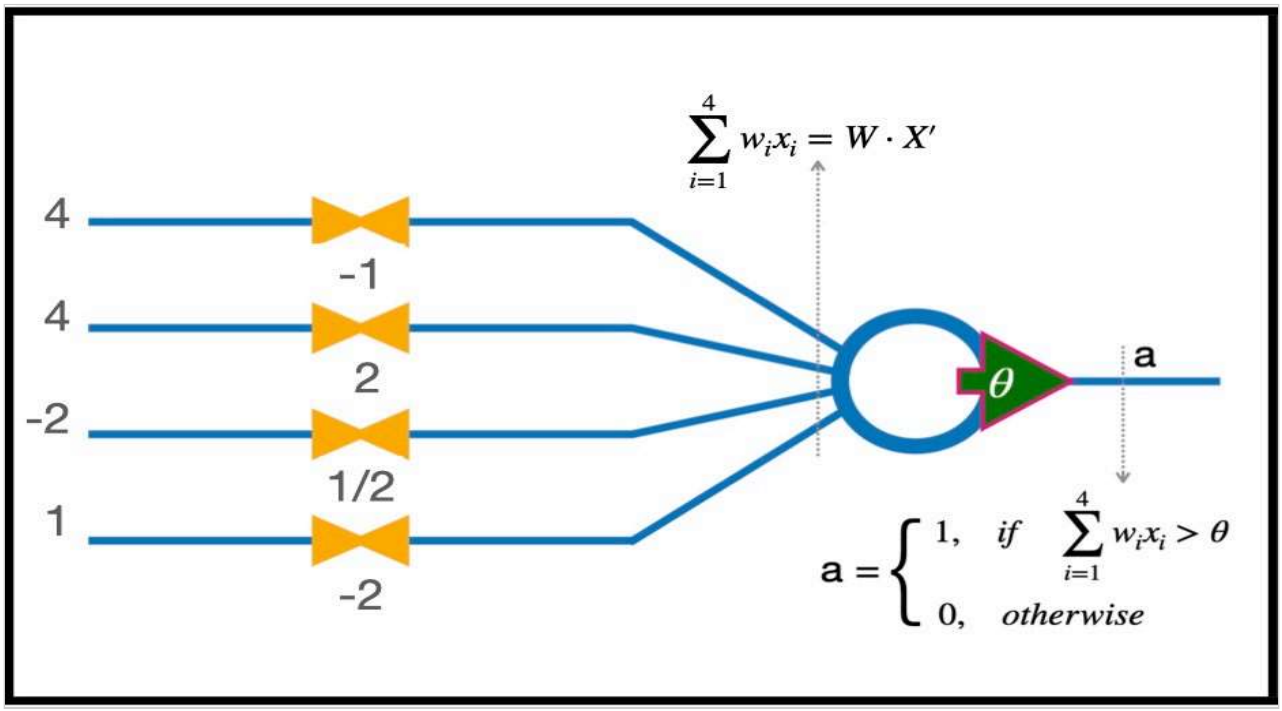
**2 * (-1)** + **3 * 2** - 2 = -2 + 6 - 2 = 2
max(0,2) = 2

**a = 2**

Ord: 23

Besvarad.

## 5 Step activation with θ




Consider the unit represented in the figure, with its input and weight distribution, step function activation, and a threshold θ = 2.
Calculate the weighted sum and its corresponding activation. (report your calculations) (**1p**)

**Fill in your answer here**

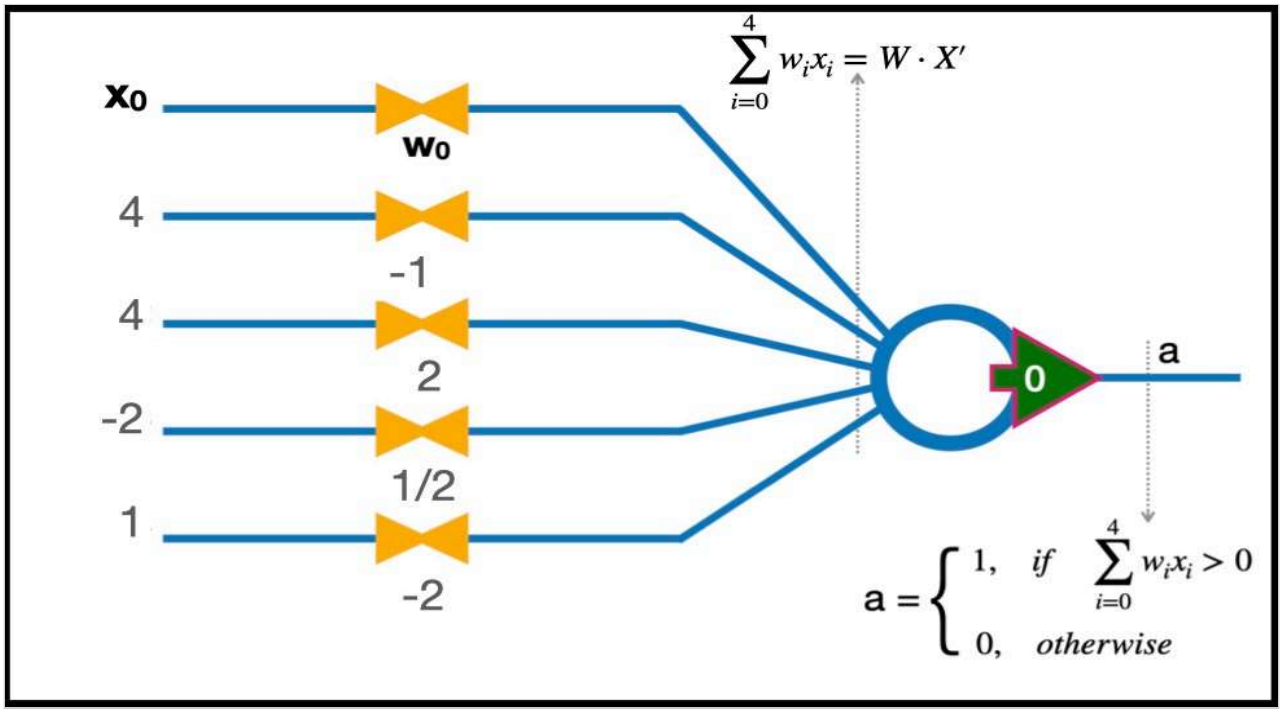
**4 * (-1)** + **4 * 2** + **(-2) * 0.5** + **1 * (-2)** - 2 = **-4** + **8** + **(-1)** + (**-2**) - 2 = -1
-1 < 0

**weighted sum = -1**
**a = 0**

Ord: 39

Besvarad.

# 6 Step activation with extra input




Consider the step function unit from the previous question, with its input and weight distribution, and a threshold θ = 2.
Complete the equivalent description of the unit by indicating the values for $x_0$ and $w_0$ (see figure above). Calculate the weighted sum and its corresponding activation. (report your calculations) (**1p**)

**Fill in your answer here**

**(-1) * (-2)** + **4 * (-1)** + **4 * 2** + **(-2) * 0.5** + **1 * (-2)** + 0 = 2 + (-4) + 8 + (-1) + (-2) + 0 = 3
3 > 0

**Weighted sum = 3**
**a = 1**

Ord: 45

Besvarad.

## 7 Activation function

A unit (artificial neuron) generates an activation value of $-1$. Which of the following activation functions are compatible with that fact? (**1p**)

**Select one or more alternatives:**

- [ ] Sign ✔
- [ ] Sigmoid
- [x] Tanh ✔
- [ ] ReLU

Delvis rätt. 0.5 av 1 poäng.

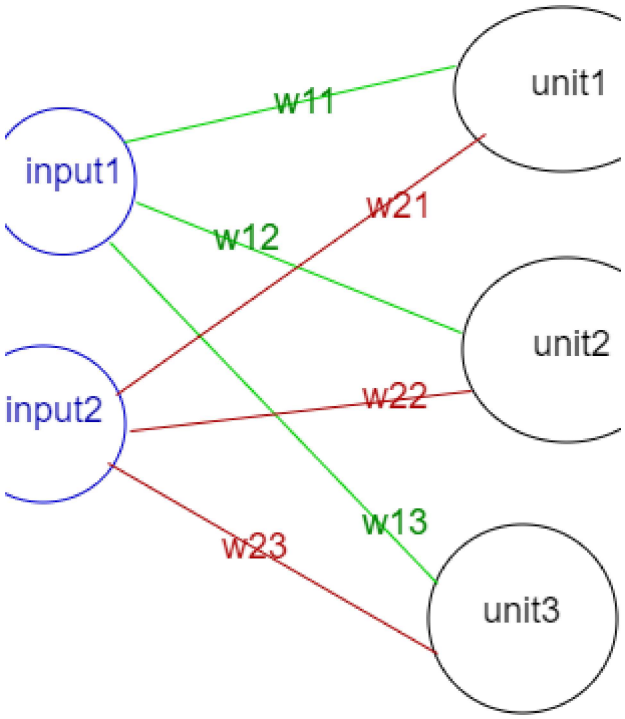
## 8 Matrix layer representation

$$\begin{pmatrix} H_{11} & H_{12} & H_{13} \\ H_{21} & H_{22} & H_{23} \end{pmatrix} = \begin{pmatrix} 4 & 5 & 2 \\ 3 & 6 & 1 \end{pmatrix}$$

The matrix of weights displayed above represents the first hidden layer of a fully connected multiple layer network.
Specify the number of inputs, the number of units in the hidden layer, and the weight value of each connection between input and the second unit of the hidden layer. (**1p**)

**Fill in your answer here**

**Two units in the input layer**, which connects to **three units in the hidden layer.**
(w11 = h11 = 4), (w12 = h12 = **5)**, (w13 = h13 = 2)
(w21 = h21 = 3), (w22 = h22 = **6)**, (w23 = h23 = 1)
Input1 connects to unit2 with h12/w12 = **5** and input2 connects with unit2 with h22/w22 = **6**




Ord: 62

Besvarad.

# 9 Layout 1

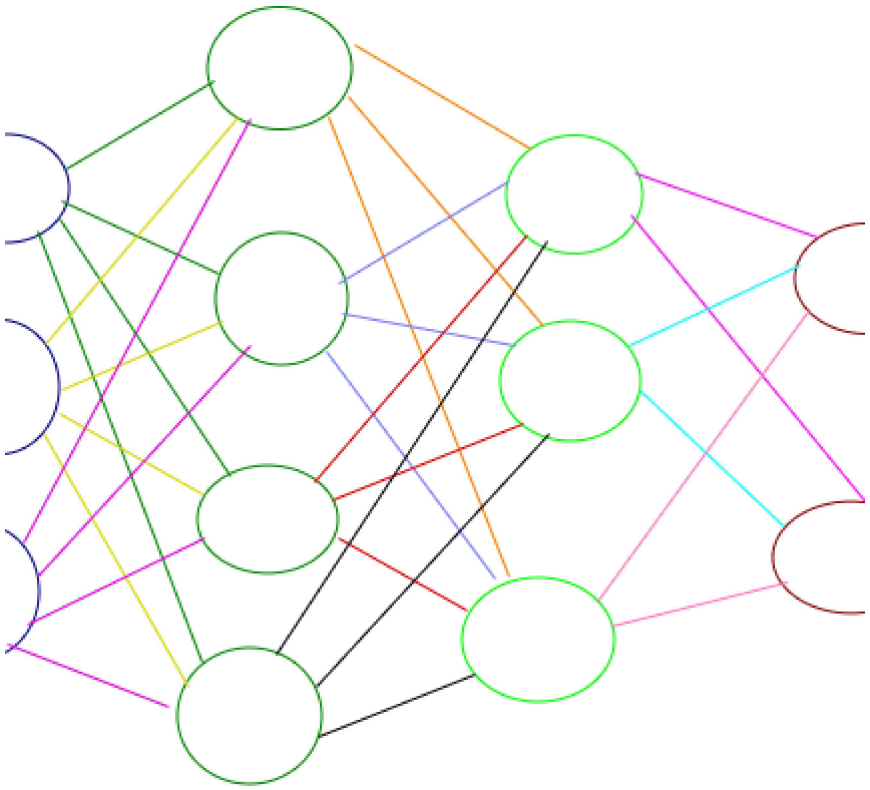
Consider a fully connected network, with three input units, a first hidden layer constituted of four units, a second hidden layer of three units, and a two unit output layer.
Calculate the total number of weights and the total number of biases. (**1p**)

**Fill in your answer here**

First hidden layer has 4 biases, second has 3, and the output has 2, 4 + 3 + 2 = 9, **9 biases total**

From input to first hidden layer 3 * 4 = 12, first hidden layer to second, 4 * 3 = 12, second to output, 3 * 2 = 6

12 + 12 + 6 = **30, 30 weights.** 30 weights + 9 biases = 39 parameters



Ord: 70

Besvarad.

## 10 Layout 2

```
model = Sequential()
model.add(Dense(10, activation='sigmoid', input_shape=(100,)))
model.add(Dense(10, activation='softmax'))
```

Consider the code snippet in the figure above.
Briefly describe the network layout. For each layer, specify the number of weights and biases. (**1p**)

**Fill in your answer here**

**100 units** in the input layer. **10 units** with sigmoid neurons in the hidden layer, with **10 biases**. 10 units with softmax activation functions in the output layer, with **10 biases.** Total of **20 biases.** Each layer is fully connected, resulting in 100 * 10 + 10 * 10 = **100 000 weights**.

**20 biases**

**100 000 weights**

Ord: 58

Besvarad.

## 11 Backpropagation 1

Is the Tanh function a suitable activation function for training by backpropagation? Why? (**1p**)

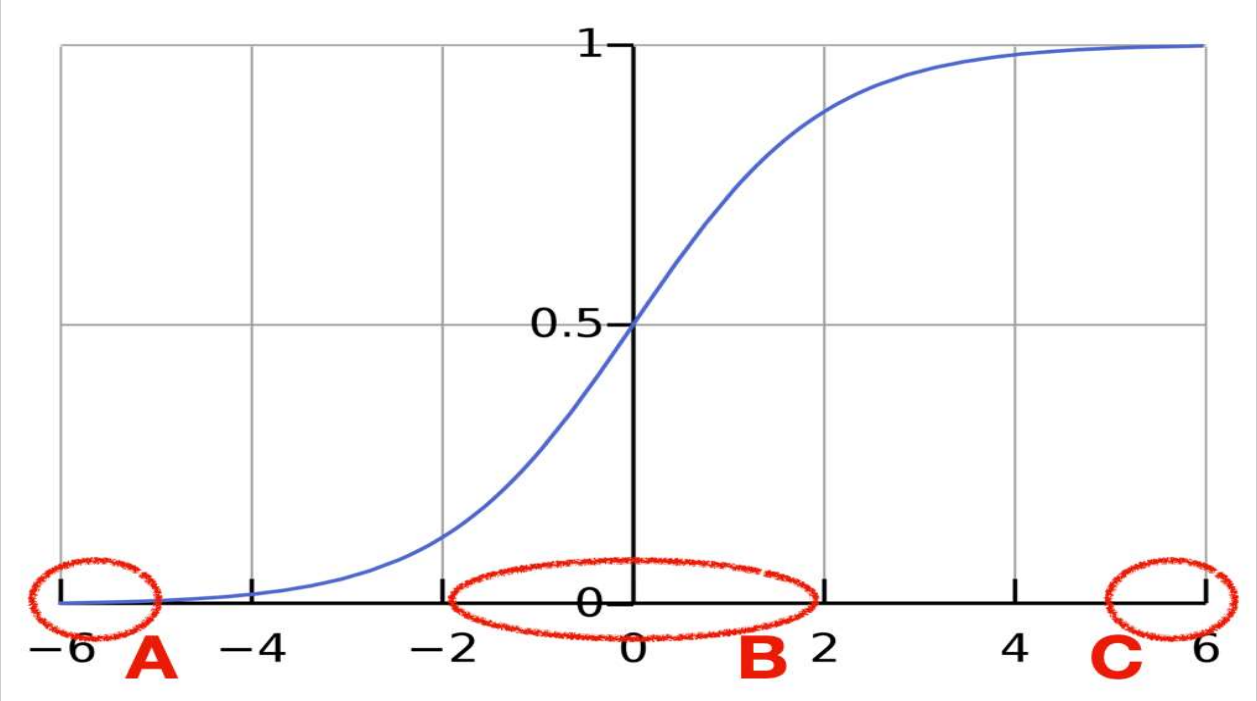
**Fill in your answer here**

Yes, the Tanh activation function produce values around x>0(positiv values, example: 0,7) if the values are positiv, around 0 if the values are near 0, and values around x<0(negativ values, example: -0,4 ) if the values are negativ.
The larger the value is, the closer the Tanh function will output something close to 1, 0,9 for example.
The smaller the value is, the closer the Tanh function will output something close to -1, -0,8 for example.
Since the differences is so small, small changes can have an impact, which makes it easier to adjust the values and thus train the network through backpropagation.

Ord: 103

Besvarad.

## 12 Backpropagation 2




Refer to the figure. During training by backpropagation, what is the ideal working region for the argument of the sigmoid activation function and why? (**1p**)

**Select one alternative:**

- C - Because of the higher value of the function.
- B - Because of the intermediate value of the function.
- C - Because of the limited value of the function's slope.
- A - Because of the lower value of the function.
- B - Because of the maximum value of the function's slope. ✓

Rätt. 1 av 1 poäng.

## 13 Learning rate

What is the effect of the learning rate η? (**1p**)
**Select one alternative:**

- To determine the size of the mini-batch.
- To determine the level of dropout.
- To control the level of update of the weights during learning. ✓
- To control the level of activation of the trained unit.

Rätt. 1 av 1 poäng.

## 14 Batches

With reference to the use of mini-batches in backpropagation, which statements are correct? (**1p**)
**Select one or more alternatives:**

- Considering each epoch, the weights are updated along the direction of the steepest gradient. ✗
- The sequence of training samples is different for each epoch. ✓
- During each epoch, each item in the training set is shown to the network once and only once. ✓
- Splitting the training data into smaller batches can result in increased memory usage by the system.

Delvis rätt. 0 av 1 poäng.

## 15 Number of neurons

Imagine you have trained a network for image classification with 64 units in a given hidden layer. Reducing the number of units in that hidden layer to 32 does not result in any reduction in the performance. Based on that observation, what is the most reasonable action to take next? (**1p**)

**Select one alternative:**

- Test the performance for 128 units.
- Test the performance for 16 units. ✓
- Test the performance after halving the value of the learning rate.

Rätt. 1 av 1 poäng.

## 16 Sat

Name the technique that can prevent saturation, particularly during the initial epochs of training a neural network using backpropagation. (**1p**)

**Fill in your answer here**

**Glorot Weight initialization**, when asigning random values to the weights in the begining, many of the values tend to very large or very small, which is bad since very large or very small values means strong opinions. It also leads to neuron saturation, since adjustment to a large value has very little impact on the value. Most neurons have their golden spot in the middle, where the slope will be the steepest, and learning will be the most efficient.

Ord: 79

Besvarad.

## 17 Vanishing gradients

Is it effective to use Glorot initialization and batch normalization simultaneously? Explain why or why not. (**2p**)

**Fill in your answer here**

Yes, glorot initialization will asign values to the weights(w) which corresponds to small z(x * w + b = z) in the beginning at the first epoch. Batch normalization will help backpropagation through out the training, by making sure that large values in on layer, won't have an large effect on the next layer. Using both might by some be considered overkill, since they both deal with large weight and bias values. But to avoid neuron saturation, at the begining aswell as through out the training, is worth it. **Yes.**

Ord: 90

Besvarad.

## 18 Convolutional and pooling layers

Consider an 6x6x3 RGB image (i.e. an 6x6 image with three color channels) with no padding. Apply one 3x3 kernel with a stride of 1. What is the size and depth of the resulting activation map?

Then, apply onto such an activation map a 2x2 max pooling layer with a stride of 2. What is the size and depth of the final map? (**2p**)

**Fill in your answer here**

Yes

Ord: 1

Besvarad.