GÖTEBORGS UNIVERSITET

STUDENT

0015-CNS

TENTAMEN

TIG122 Tentamen

Kurskod	--
Bedömningsform	DT
Starttid	18.03.2024 14:00
Sluttid	18.03.2024 17:00
Bedömningsfrist	--
PDF skapad	20.01.2025 15:39
Skapad av	Catarina Elg

i General information

- 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.
- From 15:00, the course coordinator will be present in the exam room to answer questions about the exam.
- Question types:
  - **Multiple choice (one correct answer): 7 questions**
  - **Multiple choice (multiple possible correct answers): 1 question** --> ATTENTION: selecting the wrong answer(s) zeroes the gain from that question!
  - **Free text: 3 questions** --> Show your knowledge/understanding of the issue at hand in a relatively short form.
  - **Basic calculations: 7 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 -- information cloned from "Canvas TIG122 VT24 > Syllabus" --

The Bonus points will be added to the examination score, after being transformed by the "ceiling" function, i.e. the Bonus points will be transformed into the least integer number greater than or equal to their value (e.g. 0.5 Bonus points are transformed into 1).

Total score = Examination score + ceil(Bonus points)

- Total score ≥ 14 --> VG
- 8 ≤ Total score <14 --> G
- Total score < 8 --> U

In case the Total score misses the pass threshold (i.e. Total score= 8), the Bonus points will be reconsidered during reexamination.

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?

Fill in your answer here

Ord: 0

Obesvarad.

2 Decision trees

Which statement is correct?
Select one alternative:

- The point of bootstrapping is to permute the order of the table rows.
- Decision trees is an ensemble method.
- The last step in random forest classification is voting. ✓

Rätt. 1 av 1 poäng.

3 Regression trees

Which statement is correct?
Select one alternative:

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

Fel. 0 av 1 poäng.

4 ReLU activation

Consider a unit with a ReLU activation function, input x = (4,3), weights w = (−1/2, 1), and threshold θ = 3.
Calculate its activation a. (report your calculations)

Fill in your answer here

wx+b
-threshold = b

4*-1/2 = -2

3*1 = 3

-2+3 = 1

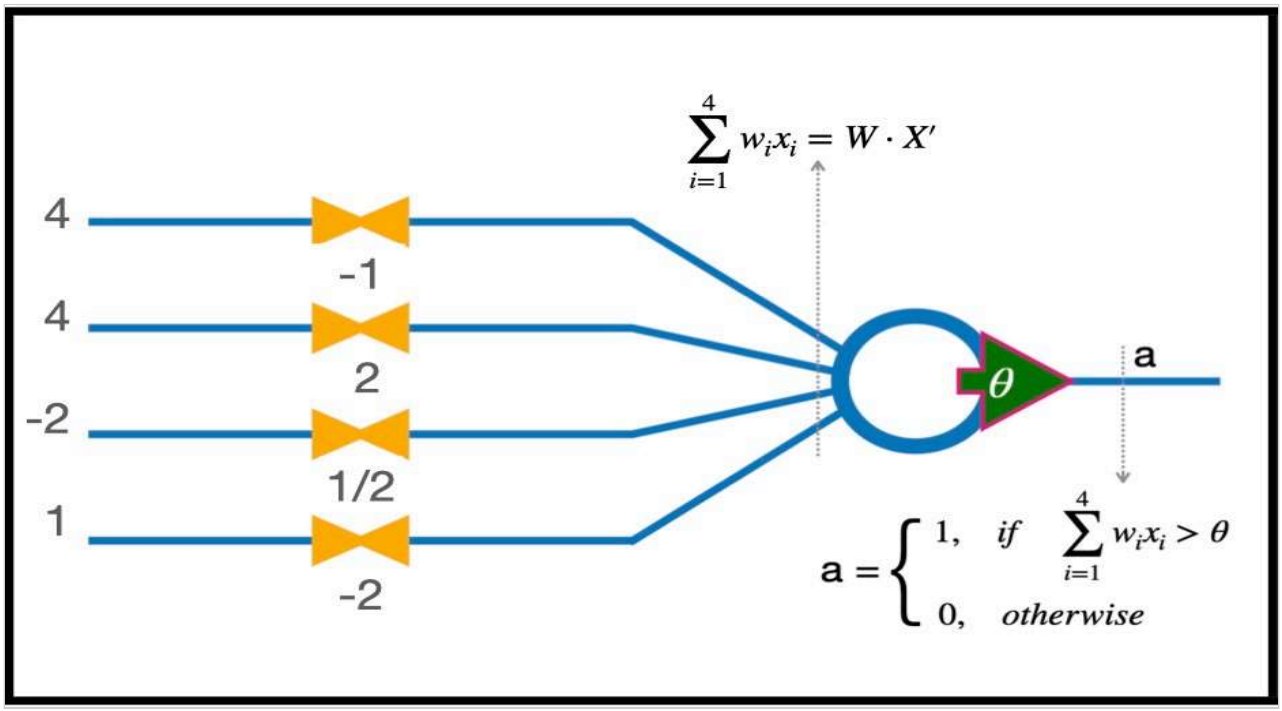
1-3 = -2

in ReLU activation function, if weighted sum under 0 the activation a is **0.**

Ord: 29

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)

Fill in your answer here

(4*-1)+(4*2)+(-2*1/2)+(1*-2) = -4 + 8 -1 -2 = **1**

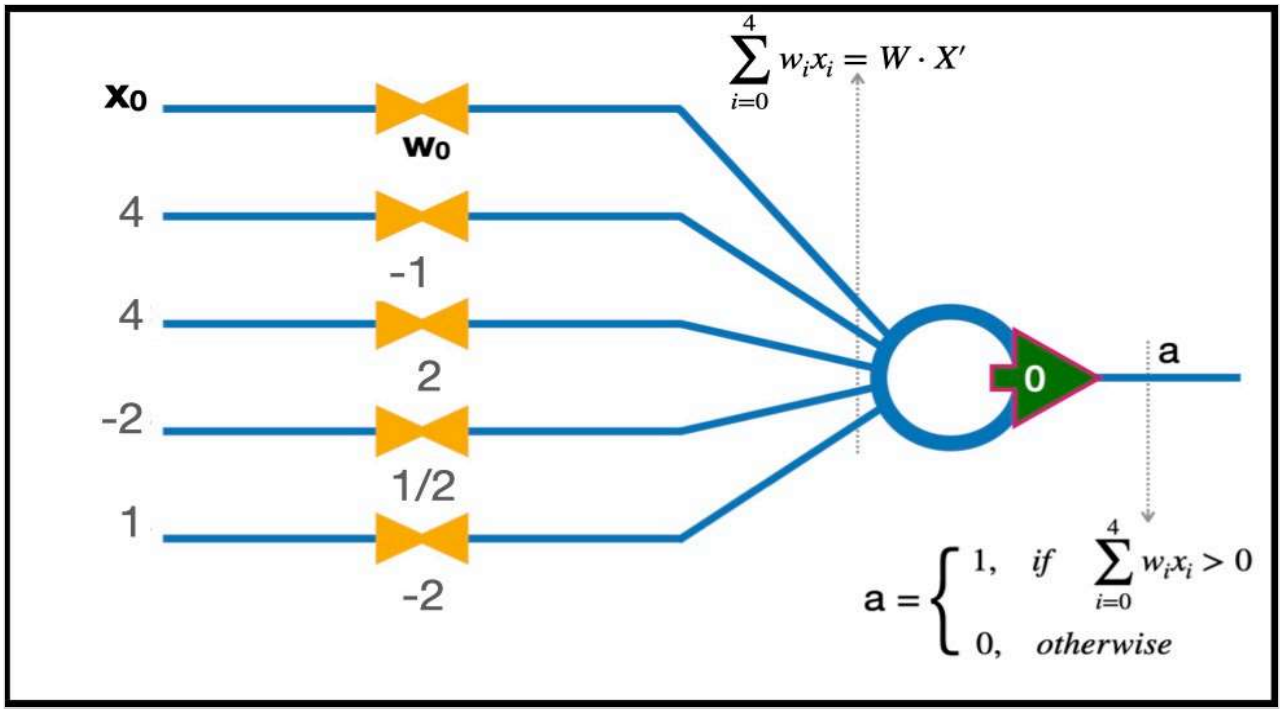
Looking at the activation function we can see that 1 is more than our threshold -2 which means activation a is 1.

Weighted sum = 1
Activation a = 1

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)

Fill in your answer here

x0 = 1
w0 = the threshold from previous question = -2

(1*-2)+(4*-1)+(4*2)+(-2*1/2)+(1*-2) = -1

As -1 is lower than 0 as per our activation function our activation a is 0.

Weighted sum = -1
Activation a = 0

Ord: 38

Besvarad.

7 Activation function

A unit (artificial neuron) generates an activation value of 2. Can its activation function be ReLU?
Select one alternative:

- Yes. ✔
- No.
- It depends on the threshold. ✖

Fel. 0 av 1 poäng.

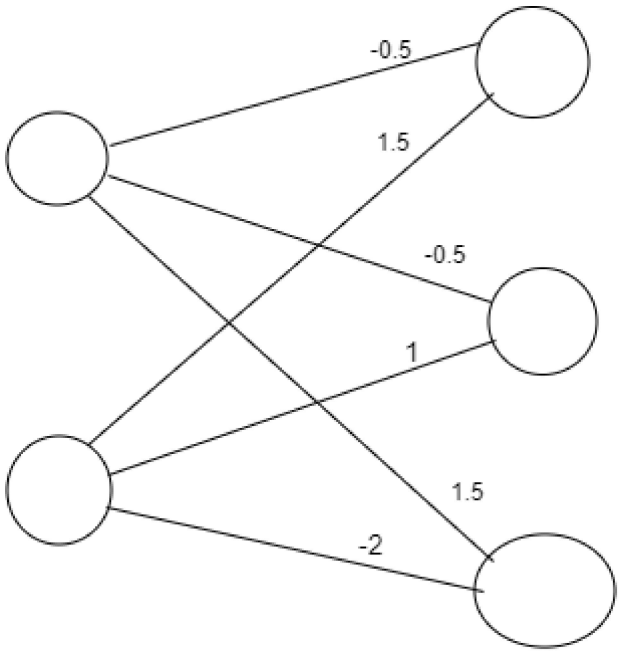
8 Matrix layer representation

$$\begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \\ H_{31} & H_{32} \end{pmatrix} = \begin{pmatrix} -0.5 & 1.5 \\ -0.5 & 1 \\ 1.5 & -2 \end{pmatrix}$$

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.

Fill in your answer here

2 inputs, 3 units in hidden layer.



I show all the connections here which I assume won't be a problem even though only connections to the second hidden unit was asked..

Ord: 31

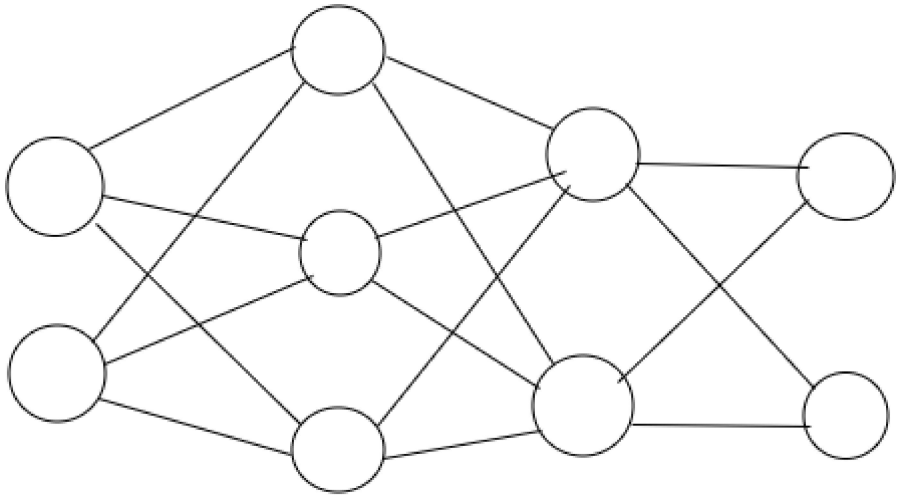
Besvarad.

9 Layout 1

Consider a fully connected network, with two inputs, a first hidden layer constituted of three units, a second hidden layer of two units, and a two unit output layer.
Calculate the total number of weights and the total number of biases.

Fill in your answer here

Every connection has a weight. 6+6+4 = 16 weights.
Every unit except the input layer has a bias. 3+2+2= 7 biases.



Ord: 21

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.

Fill in your answer here

The model is initialized with a sequential model.

It then adds its first input layer as well as the first hidden fully connected layer with sigmoid activation.
After that it adds its output layer with using softmax.

The input layer has 100 units, the first hidden layer 10 units and the output layer 10 units.

weights = (100*10)+(10*10) = 1100 weights.
biases = 10+10 = 20 (again the input layer does not have biases.)

Ord: 74

Besvarad.

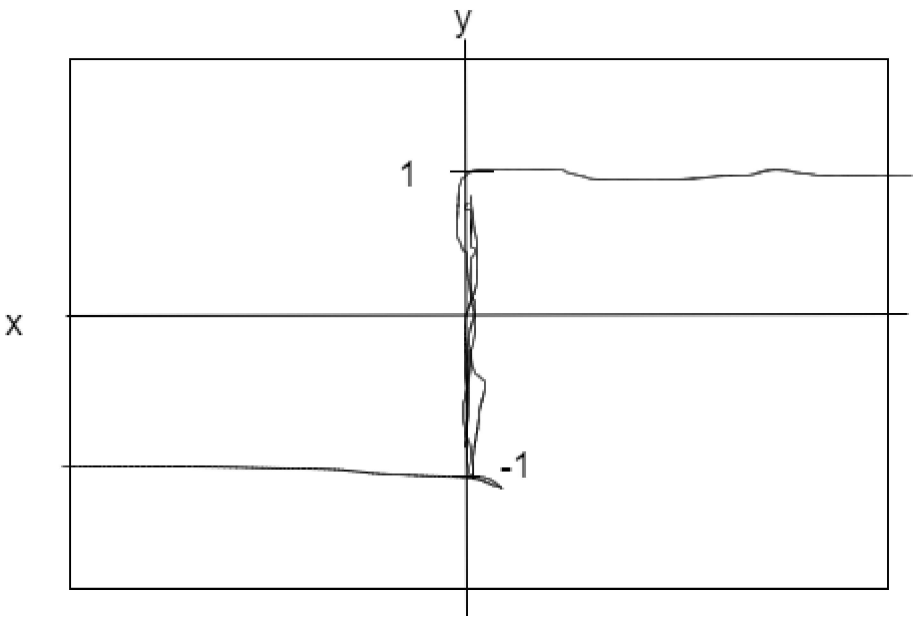
11 Backpropagation 1

Is the sign function a suitable activation function for training by backpropagation? Why?

Fill in your answer here

It is not. sign function has its activation as either 1 or -1, no in between.

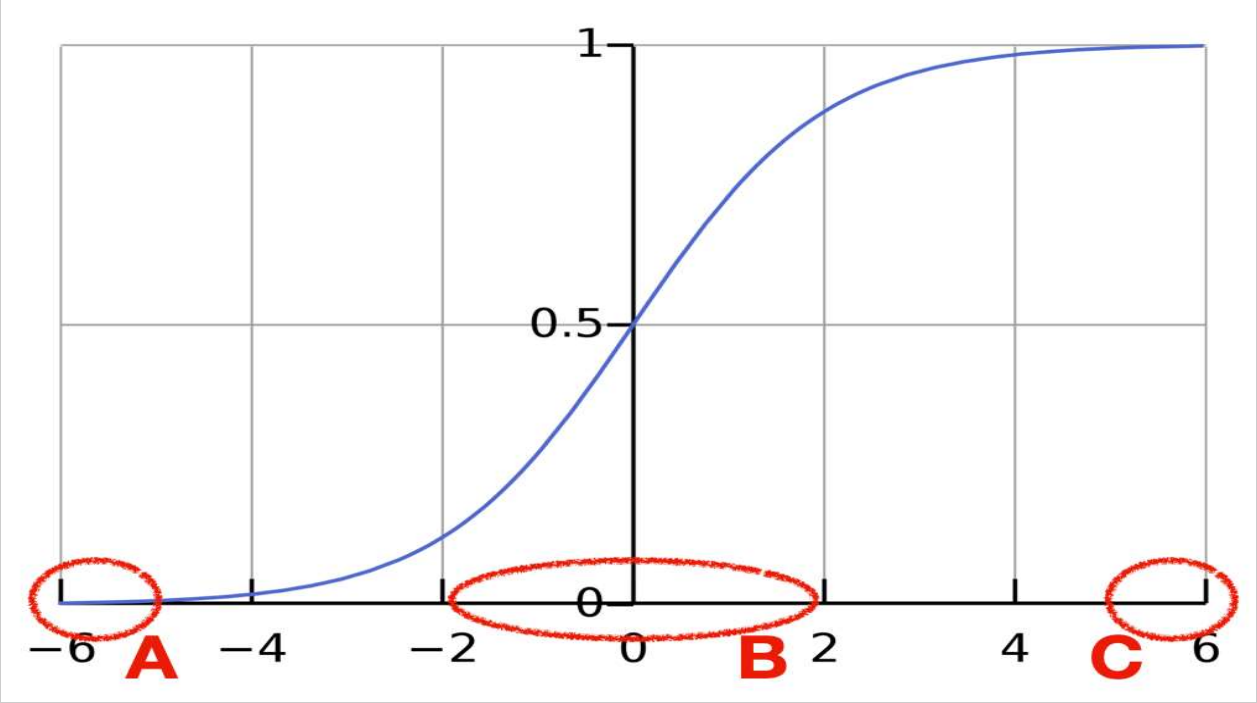
As backpropagation uses the gradient to change the models weights and biases it's not flexible at all compared to something like sigmoid where the activation can be more "granular" values to update the weights and biases with more "accurate" values. To clarify, the gradient for sign activation function don't really have many options as it looks like this:



Ord: 74

Besvarad.

12 Backpropagation 2



Refer to the figure. During training by backpropagation, what is the problem related to the argument of the sigmoid activation function falling in region A?

Select one alternative:

- Negative value of the function's argument.
- Low value of the function's slope. ✓
- Low activation value.

Rätt. 1 av 1 poäng.

13 Learning rate

What is the effect of the learning rate η?
Select one alternative:

- (●) To stabilize the process of weight convergence. ✓
- ( ) To prevent overfitting.
- ( ) To slow down learning.

Rätt. 1 av 1 poäng.

14 Batches

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

- [ ] Splitting the training data into smaller batches can result in increased memory usage by the system.
- [ ] During each epoch, each item in the training set is shown to the network once and only once. ✓
- [x] The sequence of training samples is different for each epoch. ✓

Delvis rätt. 0.5 av 1 poäng.

15 Number of layers

Imagine you have trained a network for image classification with three hidden layers. Increasing the number of layers to four does not result in any increment in the performance. Based on that observation, what is the most reasonable action to take next?

Select one alternative:

- Test the performance for two hidden layers. ✓
- Test the performance after doubling the number of units per layer.
- Test the performance for five hidden layers.

Rätt. 1 av 1 poäng.

16 Dropout

After training a neural network while using dropout to facilitate model generalization, which measures must be taken during validation or use?

Select one alternative:

- Multiply the weights at each layer by a factor that is consistent with the dropout rate at that layer. ✓
- Divide the weights at each layer by a factor that is consistent with the average dropout rate.
- Apply the input and obtain the output as if dropout was not used.

Rätt. 1 av 1 poäng.

17 Vanishing gradients

What is the origin of the problem known as ”vanishing gradients”? Is it more or less severe in deeper networks, i.e. networks with a higher number of layers? Name (at least) one possible solution to the problem and briefly explain why it can be effective.

Fill in your answer here

Vanishing gradients happens when the learning slows down drastically because the gradient is way too small on the curve while for example using sigmoid. As you often multiply with the gradient in backpropagation, if the gradient has a very small value it will barely update the weights therefore stagnate the learning. It can be caused by badly initialized weights and is more common in activation functions like sigmoid and tanh.

It it more severe in deeper networks with more layers as backpropagation in the backwards step happens layer by layer and therefore the updates gets worse and worse for each layer there is.

One possible solution to vanishing gradients is L1/L2 regularization that adds a parameter to the calucations to prevent too big/small/complex weights.

Ord: 124

Besvarad.

18 Convolutional and pooling layers

Consider an 8x8x3 RGB image (i.e. an 8x8 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?

Fill in your answer here

The resulting activation map has a size of 6x6 and a depth of 6x6x3 = 108.
The final map has a size of 3x3 and a depth of 3x3x3 = 27.

Ord: 31

Besvarad.