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

Format | Σ

Words: 0

Maximum marks: 1

2 Decision trees

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

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

Maximum marks: 1

3 Regression trees

Which statement is correct? (**1p**)

Select one alternative:

- 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.
- When constructing regression trees, the goal is to minimize the Gini impurity.

Maximum marks: 1

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**)

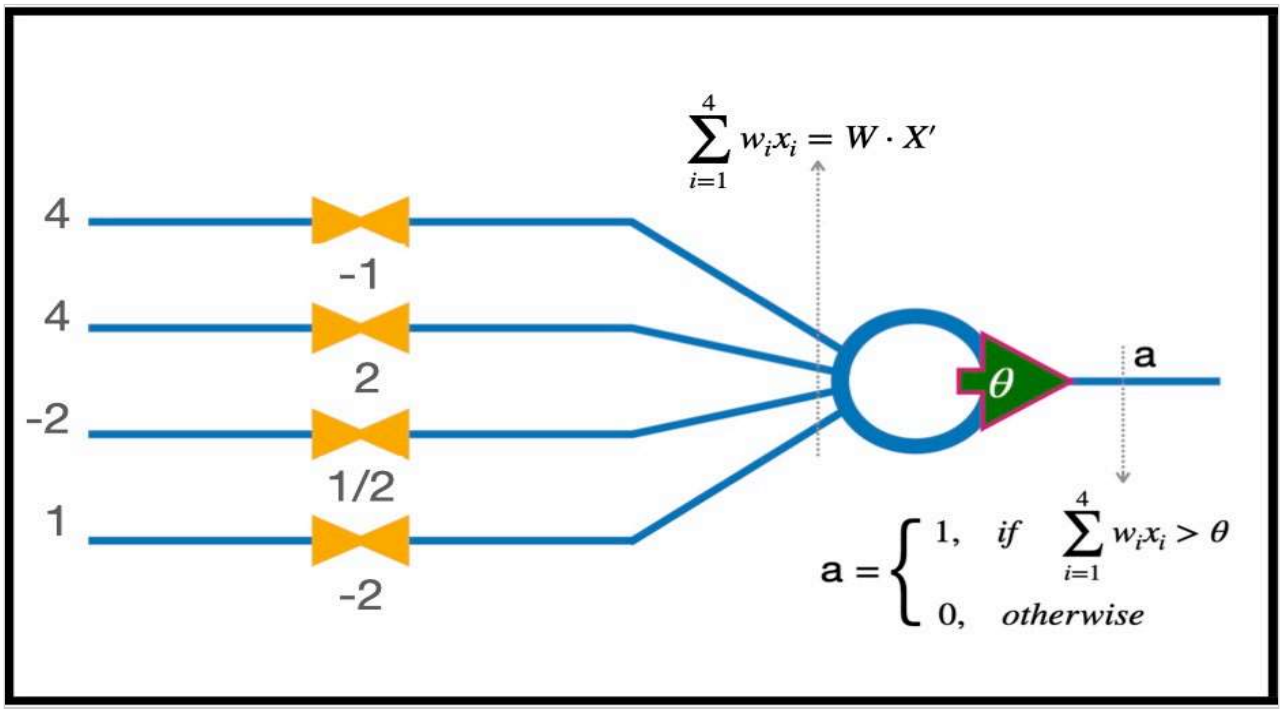
Fill in your answer here

Format | Σ

Words: 0

Maximum marks: 1

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**)

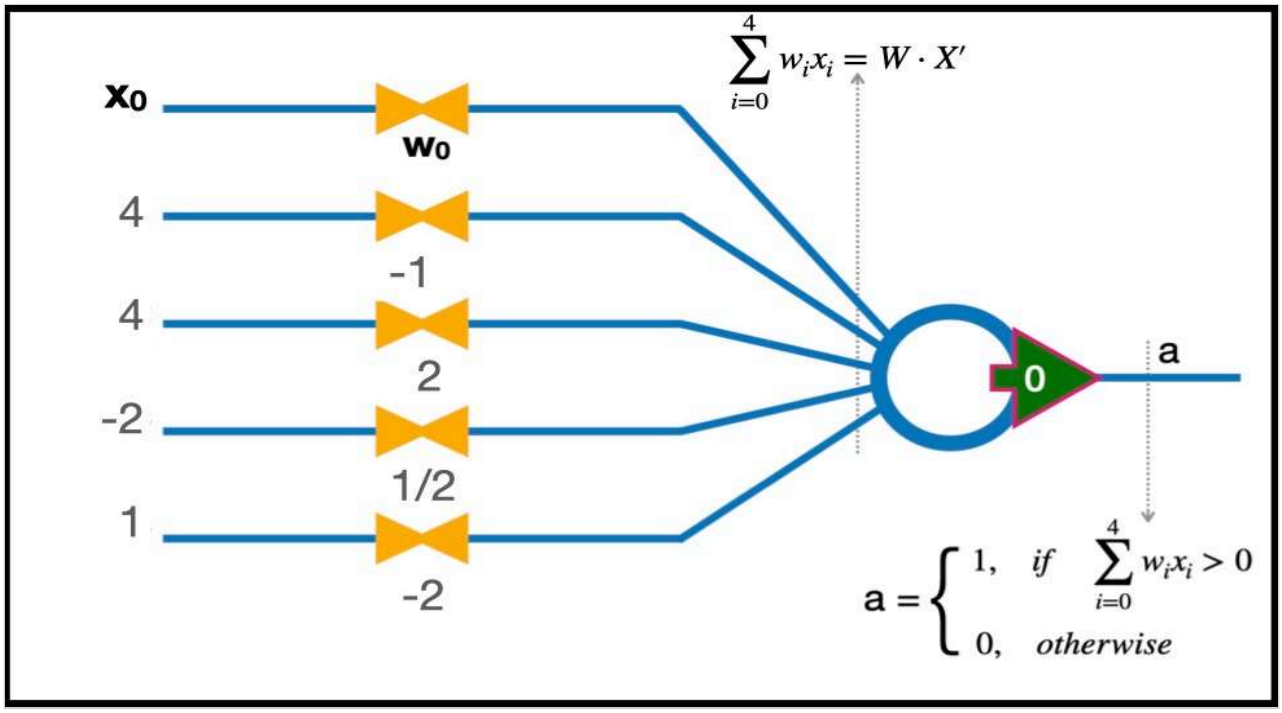
Fill in your answer here

Format

Words: 0

Maximum marks: 1

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

Format

Words: 0

Maximum marks: 1

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:

- [ ] Tanh
- [ ] Sigmoid
- [ ] ReLU
- [ ] Sign

Maximum marks: 1

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

Format | | | | | | Σ |

Words: 0

Maximum marks: 1

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

Format | Σ

Words: 0

Maximum marks: 1

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

Format

Words: 0

Maximum marks: 1

11 Backpropagation 1

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

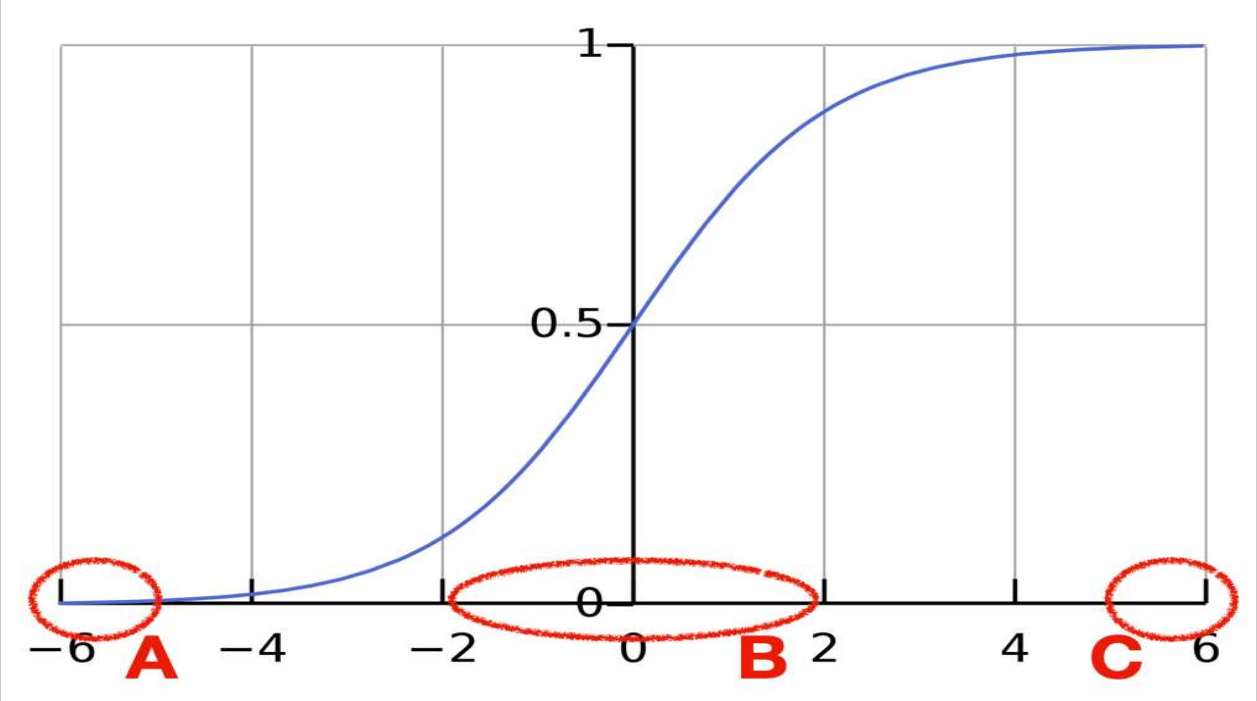
Fill in your answer here

Format

Words: 0

Maximum marks: 1

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:

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

Maximum marks: 1

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

Maximum marks: 1

14 Batches

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

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

Maximum marks: 1

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

Maximum marks: 1

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

Format | | | | | | Σ |

Words: 0

Maximum marks: 1

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

Format | | | | | | Σ |

Words: 0

Maximum marks: 2

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

Format

Words: 0

Maximum marks: 2