![Göteborgs Universitet logo]

STUDENT

001ABC-0038-EYT

TENTAMEN

TIG117 Tentamen

Kurskod	--
Bedömningsform	DT
Starttid	21.03.2022 08:30
Sluttid	21.03.2022 11:30
Bedömningsfrist	--
PDF skapad	17.11.2022 17:12

Section 1: Simple Neural Networks

Fråga	Uppgiftstitel	Status	Poäng	Uppgiftstyp
1	Exam Information	Fel	0/0	Textfält
2	Q1 S1	Delvis rätt	0/2	Flersvarsfråga
3	Q2 S1	Rätt	2/2	Flervalsfråga
4	Q3 S1	Rätt	2/2	Flervalsfråga
5	Q4 S1	Rätt	2/2	Flersvarsfråga
6	Q5 S1	Rätt	4/4	Flersvarsfråga
7	Q6 S1	Rätt	2/2	Flervalsfråga
8	Q7 S1	Rätt	2/2	Flervalsfråga
9	Q8 S1	Rätt	2/2	Flervalsfråga
10	Q9 S1	Rätt	2/2	Flersvarsfråga

Section 2: Autoencoders

Fråga	Uppgiftstitel	Status	Poäng	Uppgiftstyp
11	Q1 S2	Obesvarad	0/2	Flervalsfråga
12	Q2 S2	Obesvarad	0/2	Flersvarsfråga
13	Q3 S2	Obesvarad	0/2	Flervalsfråga
14	Q4 S2	Obesvarad	0/2	Flervalsfråga
15	Q5 S2	Obesvarad	0/2	Flervalsfråga
16	Q6 S2	Obesvarad	0/3	Flersvarsfråga
17	Q7 S2	Obesvarad	0/2	Flersvarsfråga
18	Q8 S2	Obesvarad	0/3	Flersvarsfråga
19	Q9 S2	Obesvarad	0/2	Flervalsfråga

Section 3: Convolutional Neural Networks

Fråga	Uppgiftstitel	Status	Poäng	Uppgiftstyp
20	Q1 S3	Rätt	2/2	Flervalsfråga
21	Q2 S3	Rätt	2/2	Flervalsfråga
22	Q3 S3	Delvis rätt	0/2	Flersvarsfråga
23	Q4 S3	Delvis rätt	0/2	Flersvarsfråga
24	Q5 S3	Rätt	6/6	Flervalsfråga
25	Q6 S3	Rätt	2/2	Flersvarsfråga
26	Q7 S3	Delvis rätt	0/2	Flersvarsfråga
27	Q8 S3	Rätt	2/2	Flersvarsfråga

Section 4: Training Neural Networks and Evaluating Performance

Fråga	Uppgiftstitel	Status	Poäng	Uppgiftstyp
28	Q1 S4	Delvis rätt	0/2	Flersvarsfråga
29	Q2 S4	Rätt	2/2	Flervalsfråga
30	Q3 S4	Rätt	2/2	Flersvarsfråga
31	Q4 S4	Fel	0/2	Flervalsfråga
32	Q5 S4	Rätt	2/2	Flervalsfråga
33	Q6 S4	Rätt	2/2	Flervalsfråga
34	Q7 S4	Rätt	2/2	Flervalsfråga
35	Q8 S4	Delvis rätt	0/2	Flersvarsfråga
36	Q9 S4	Fel	0/2	Flervalsfråga
37	Q10 S4	Rätt	2/2	Flersvarsfråga

1 Exam Information

TIG117: *Machine Learning & Data Analysis*
Final Exam, March 21st 2022

Answer <u>all</u> questions of <u>any three</u> of SECTIONS 1, 2, 3 and 4.

Therefore, total sections answered must be 3!

Total available points = 60p (20 per Section)

To obtain G = 30p (50%). To obtain VG = 45p (75%).

The number of questions may vary per section but the total available points is the same per section.

All questions are multiple choice.

Questions require either one answer only or multiple answers.

Where multiple answers are required, it is necessary to get the answer completely correct to obtain the marks (it is all or nothing!).

Please, when you have chosen the sections you wish to complete, put the numbers of the

sections (with space or comma between) here: | 1,3,4 | ❌ (uppgift)

Only 3 sections will be marked. You can pass over questions on the one section you have not selected.

Totalpoäng: 0

2   Q1 S1

Below is a truth table for a simple classification problem.

Object	Purple?	Round?	Eat?
Blueberry	1	1	1
Golf ball	0	1	0
Violet	1	0	0
Hot Dog	0	0	1

Four labelled objects are required to be classified as eatable/edible or not. The ANN receives as inputs a two-element vector of the object properties "purple", "round" and must make the decision ("eat" classification) based on those properties. Which of the following are true:

More than one correct answer:

☑ The problem requires one (or more) hidden layers to solve.   ✅

☑ The input-output mapping is not linearly separable.   ✅

☐ A single layered perceptron with two input nodes and one output node could solve this problem.

☐ The problem can be viewed as a (Not) XOR problem (inverse of XOR problem).   ✔

Totalpoäng: 2

³ Q2 S1

How many weights (excluding threshold/bias weights) does a fully-connected/dense ANN with a single hidden layer with 4 input nodes, 3 hidden nodes, and 1 output node, have?
Only one of the below is correct:

○ 15 ✅

○ 12

○ 16

○ 8

Totalpoäng: 2

⁴ Q3 S1

Given an input vector of **x** = [x1,x2] where x1 = 0.5, x2 = 1.0, and a weight vector **w** = [1.0, 0.5], what is the activation value as calculated by the dot product of **x** and **w**.
One correct answer:

○ 0.25

○ [1.5, 1.5]

○ [0.5, 0.5]

○ 1.0 ✅

Totalpoäng: 2

5 Q4 S1

Which of the following are properties of the sigmoid function?
More than one correct answer

- ☑ Output can be seen as a biological analogy to probabilistic firing (of a neuron) ✅
- ☑ Output values are between 0 and 1 ✅
- ☐ Output values are between -1 and 1
- ☐ Input values must be positive

Totalpoäng: 2

6   Q5 S1

The extended truth table below shows a possible solution for an XOR problem using a Multi-layered Perceptron (MLP) with 2 input nodes and 2 hidden nodes.

| x1	x2	h1	h2	Y
0	0	0	0	0
0	1	1	0	1
1	0	1	0	1
1	1	1	1	0

Below are a number of MLPs with different weight values for connections (including for bias nodes whose values are not included in the truth table). Which of the alternative MLPs can solve the XOR problem **according to the truth table**. Note, that the transfer function we use here is a step function where the output of a node is 1 if the activation (sum of the weighted inputs) is greater than 0, otherwise the output of the node is 0.
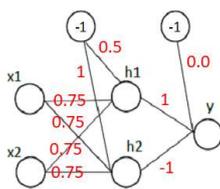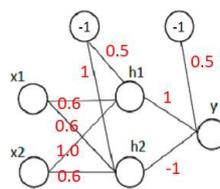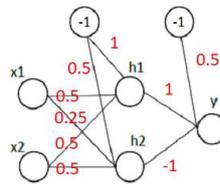

MLP A   MLP B
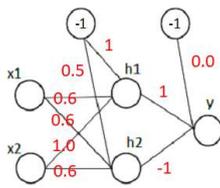
MLP C   MLP D

Choose more than one option

☐ MLP C

☑ MLP B   ✅

☐ MLP D

☑ MLP A   ✅

Totalpoäng: 4

7 Q6 S1

Which of the below best characterizes supervised learning?
Choose one of the below

○ Learning that compares ANN predictions about data to non-human labelled data

○ Learning that is based on whether or not the output of the network gives a value (or values) of 1.

○ Learning that is based on differences between input values and output values

◉ Learning that compares ANN predictions about data to human labelled data   ✅

Totalpoäng: 2

8 Q7 S1

Which of the following is most typically used as a loss/cost function for classification where there are more than 2 classes of data?
Choose one from below

○ Half mean squared classification error

◉ Cross entropy   ✅

○ Reconstruction error

○ Softmax

Totalpoäng: 2

9 # Q8 S1

In an MLP that consists of 2 inputs nodes, 2 hidden nodes and 1 output node, how many 2-input SLPs are effectively embedded within this network? (ignore bias nodes)

Choose one from below

- ◉ 3 ✅

- ○ 2

- ○ 0

- ○ 5

Totalpoäng: 2

10 Q9 S1

Below is a truth table for a simple classification problem. Purple (x1) and Round (x2) are properties of the objects (inputs). Eat (y) is the output decision.

Object	Input Vector	x1	x2	y
Blueberry		1	1	1
Golf ball		0	1	0
Violet		1	0	0
Hot Dog		0	0	0

Which of the following are true?
Choose more than one from below

☐ The truth table presents an OR problem

☑ The problem does not require one or more hidden layers for an ANN to solve it. ✓

☑ The truth table presents an AND problem ✓

☑ A single layered perceptron can find a straight line decision boundary in the input (x) to output (y) geometric mapping ✓

Totalpoäng: 2

11 Q1 S2

How many weights does a fully connected autoencoder (one hidden layer) have with 5 inputs and a bottleneck of 2 neurons/nodes (used for simple reconstruction, i.e. not classification)?
Choose one from below

- ⚪ 10

- ⚪ 20 ✔

- ⚪ 7

- ⚪ 14

Totalpoäng: 2

12 Q2 S2

Which of the below properties **are specific to** autoencoders (as compared to other types of ANNs)?
Choose more than one of the below

- ☐ They are trained by backpropagation.

- ☐ They have multiple hidden layers.

- ☐ For at least part of the training they are trained on reconstruction error. ✔

- ☐ The output layer is the same size (number of neurons) as the input layer. ✔

Totalpoäng: 2

13 Q3 S2

Why does classification error typically increase in an autoencoder with 2 hidden layers as compared to 1 hidden layer following layer-by-layer pre-training?
Choose one of the below

○ The second layer of weights are trained on imperfect data representations from the first layer. ✓

○ There are not enough data samples available to train the second layer of weights.

○ The second layer of weights are harder to train than the first layer of weights because there are more weights in this layer.

○ The second layer of weights suffer from the vanishing gradient problem

Totalpoäng: 2

14 Q4 S2

Why do deeper autoencoders (multiple hidden layers) typically outperform (according to the loss function) more shallow autoencoders (following full training)?
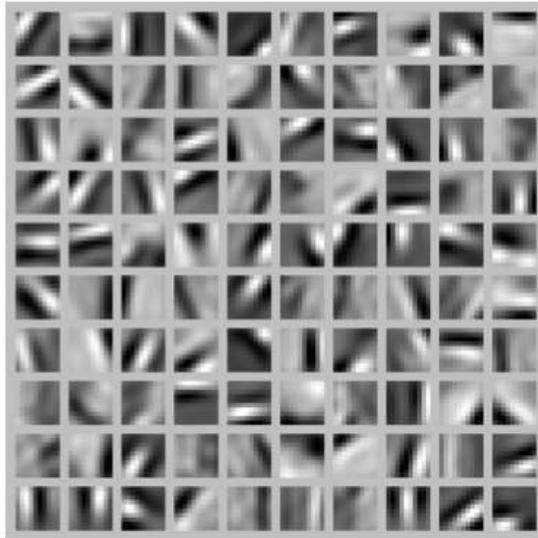**Choose one from the below**

○ Deeper autoencoders break down the learning problem into simpler features ✓

○ Deeper autoencoders have more parameters (weights)

○ Deeper autoencoders process data more quickly than shallower autoencoders

○ Deeper autoencoders have fewer parameters (weights)

Totalpoäng: 2

15 Q5 S2

The image below shows visualizations of a number of hidden nodes (inside the grey squares) in an autoencoder.

What do the individual pixels within the squares visualize?

Choose one from below

○ Activations of the hidden nodes

○ Values of the hidden nodes one they have passed through an activation function, e.g. sigmoid.

○ Weights from each input node to the specific hidden node ✔

○ Activations of the input nodes

Totalpoäng: 2

16 Q6 S2

The images below show: (left) original data (set of images) on which an autoencoder is trained on, (centre) a noisy/corrupted set of images that the trained autoencoder is presented, (right) the reconstructed inputs of the trained autoencoder.
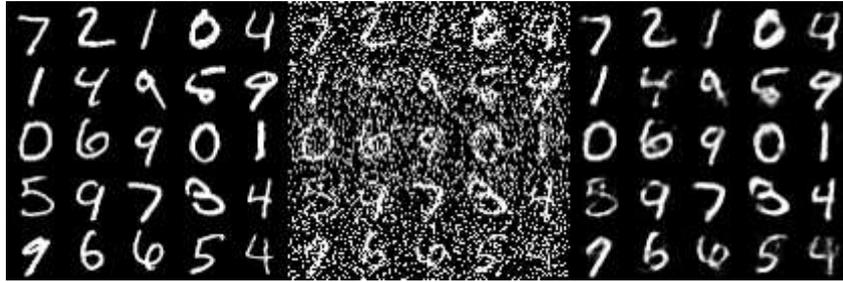
Which of the following statements are true?

Choose more than one from below

- ☐ Autoencoders trained on complex images with a large number of classes carry out lossy compression. ✓

- ☐ The autoencoder can "denoise" the corrupted data by reconstructing the noise in the input at the output layer.

- ☐ Adding a sparsity constraint in the loss function makes the autoencoder more robust to noise ✓

- ☐ Adding noise during training would mean that the trained autoencoder will output noisy images following training (even if the images presented after training are not noisy)

- ☐ The bottleneck is the part of the autoencoder where activations will typically be most affected by noisy data.

- ☐ The autoencoder can "denoise" the corrupted data as the low dimensional bottleneck filters out the values in the image that were not frequently presented during training ✓

Totalpoäng: 3

17 Q7 S2

Which of the following **activation** functions can never output a value of greater than 1 for a specific neuron/node?
Choose more than one of the below

☐ ReLU function

☐ Sigmoid function   ✔

☐ Softmax function   ✔

☐ Step/Threshold function   ✔

Totalpoäng: 2

18 Q8 S2

Transfer learning allows an ANN, including autoencoders, to be trained for one problem (e.g. classifying hand-written digits) and to utilize parts of the trained ANN for an untrained ANN on another related problem (e.g. classifying hand-written alphabetic letters). Which of the following are true for the transfer learning procedure typically used for these ANNs?
Choose more than one of the below

☐ You typically continue training the weights that are transferred from the trained ANN to the untrained ANN.   ✔

☐ Weights transfer is most commonly used for shallow (one hidden layer) autoencoders rather than deep autoencoders.

☐ Transfer learning is not a possible approach when using pre-training for stacked autoencoders.

☐ Transfer learning works by reusing the learned simple feature representations of the data relevant to one task for the new but related task.   ✔

☐ Transfer learning is more successful when the only weights of the trained network that are 'transferred' are from a layer close to the output layer.

☐ Transfer learning can make learning for the new ANN faster for the related task.   ✔

Totalpoäng: 3

19 Q9 S2

A way to view feature representation in an autoencoder is that a node/neuron is sensitive to input values that reliably correlate with its own activity. Below is a table showing a vector **x** of input elements whose values reflect 4 different data samples. These values provide inputs to a node y (weight values not shown).

x1	X2	x3	x4	x5	x6	x7	x8	x9	y
1	0	0	0	0	0	0	1	0	1
0	1	1	0	1	1	1	0	1	0
1	0	0	1	0	0	1	1	1	1
0	1	1	0	1	1	0	0	0	0

Which of the following is the complete set of full or partial features?
Choose one of the below

○ x1, x2, x3, x4, x5, x6,x8,   ✔

○ x1, x4, x8

○ x1, x2, x3, x5, x6, x8

○ x1,x8

Totalpoäng: 2

20   Q1 S3

How many pooling maps does a standard 'shallow' Convolutional Neural Network (1 layer of feature maps + 1 layer of pooling maps) have that uses 6 filters of dimensionality 4x4 over the input image?

Choose one from the below

- ○ 6   ✅
- ○ 96
- ○ 16
- ○ 22

Totalpoäng: 2

21   Q2 S3

How many neurons does a feature map in the 1st layer of a convolutional neural network have that is tied to a 2x2 filter covering a 5x5 image and that convolves over the image with stride = 1. Assume that feature map dimensions = feature map neurons.

Choose one from the below

- ○ 29
- ○ 25
- ○ 16   ✅
- ○ 4

Totalpoäng: 2

22 Q3 S3

Which of the following make CNNs efficient with respect to the number of parameters they have as compared to "dense" ANNs?

Choose more than one from below

- ☐ Adding convolutional layers to CNN always means having fewer parameters (weights) in the full network

- ☐ Deeper CNNs allow for fewer weights between the final layer and the fully connected layer (as compared to shallow CNNs) ✓

- ☑ Each feature map consists of a number of neurons whose encodings share a single filter of weights ✓

- ☑ Deeper CNNs downsample over layers ✓

Totalpoäng: 2

23 Q4 S3

Which of the following correctly describe similarities and differences between convolutional filters (encoding the feature maps) and pooling "filters" (encoding the pooling maps)?

Choose more than one from below

- ☐ Both types of filter have trainable weights

- ☐ Pooling "filters" are typically of lower dimensionality than convolutional filters ✓

- ☑ Both types of filter carry out convolutions over their processed "inputs" (e.g. image or preceding feature map layer) ✗

- ☑ Both pooling and convolutional filters allow for downsampling ✓

Totalpoäng: 2

24 Q5 S3

How many **tunable weights layer by layer**, does a Convolutional Neural Network with *one convolutional layer* have consisting of the below-listed aspects? Each filtered input to each feature map has the same receptive field. Consider feature and pooling map units as neurons.

- Input image size 7x7
- 4 input convolution filters of 2x2 dimensions
- Feature map stride 1
- Pooling 'filter' dimension 2x2
- Pooling map stride 2
- Classes = 3

Hints:

- **Consider, for calculation purposes, that you have i) weights in the feature extractor part of the network, and ii) weights from the extractor part to the classification part of the network**
- **Account for a single fully connected layer (one neuron per class) in your calculation of weights, but not the softmax layer!**

Välj ett alternativ:

○ 36 (feature extractor side) + 108 (pooling to fully connected)

○ 196 (feature extractor side) + 16 (pooling to fully connected)

○ 16 (feature extractor side) + 48 (pooling to fully connected)

○ 196 (feature extractor side) + 19 (pooling to fully connected)

○ 36 (feature extractor side) + 48 (pooling to fully connected)

◉ 16 (feature extractor side) + 108 (pooling to fully connected) ✅

Totalpoäng: 6

25 Q6 S3

Which of the following best describes "translation invariance"?
Choose more than one from below

☑ Neurons in deeper layers of the CNN are relatively insensitive to shifts of the (horizontal or vertical) position of the classifiable object in the image ✅

☑ Neurons in deeper layers are insensitive to changes in position of a feature if it does not move outside their total receptive field ✅

☑ Neurons in early layers are relatively insensitive to changes in position of a feature as long as it does not move outside the receptive field of their filter ✅

☐ Neurons in deeper layers of the CNN are not affected by full rotations of the classifiable objects in the image

Totalpoäng: 2

26 Q7 S3

What does the "DeepDream" algorithm help to visualize in CNNs?
Choose more than one from below

☑ Weights representations of features ❌

☐ Precise re-presentations of features in the data

☑ Representations of features relevant to the task (e.g. classification) ✅

☑ Neurons' sensitivity to particular features in the data ✅

Totalpoäng: 2

27 Q8 S3

What are the main differences in the way in which CNNs are applied to Generative Adversarial Networks (GANs) when processing images for the discriminator and generator networks?
Choose more than one from below

- ☑ The Discriminator downsamples; the Generator upsamples ✅

- ☑ The Discriminator must correctly predict whether data presented to it is real or fake ✅

- ☐ The Discriminator uses random noise for producing fakes

- ☐ The Discriminator upsamples; the Generator downsamples

Totalpoäng: 2

28 Q1 S4

Which of the following are considered stopping criteria when training ANNs?
Choose more than one from below

- ☐ Training loss/error below a pre-set threshold ✔

- ☑ Validation loss/error below a pre-set threshold ❌

- ☑ Successive non decreasing loss/error for validation data checks ✅

- ☑ When training runs beyond a pre-set number of epochs ✅

Totalpoäng: 2

29 Q2 S4

Neural network training data is typically split up into *mini*-batches for training. A mini-batch represents a subset of all the data available and weights are updated only once all the data in the mini-batch has been presented to the network. If the total amount of data available for training is of size 30, where mini-batches of size 5 are used for training, given a single epoch of training, how many **iterations** of training will have occurred?

Choose one from below

○ 6 ✅

○ 150

○ 30

○ 5

Totalpoäng: 2

30 Q3 S4

Which of the following are examples of hyperparameters?
Choose more than one from below

☐ Connection weights

☑ Number of hidden nodes ✅

☑ Number of hidden layers ✅

☑ Learning rate ✅

Totalpoäng: 2

31 Q4 S4

Of the following, what is the **best** characterization of the "Credit Assignment Problem"?
Choose one from below

- ⦿ How to know which nodes/neurons are responsible for the network error (and by how much) for an ANN with hidden layers ✖

- ○ How to know which weights are responsible for the network error (and by how much) for a Single Layer Perceptron

- ○ How to know which weights are responsible for the network error (and by how much) for an ANN with hidden layers ✔

- ○ How to know which nodes/neurons are responsible for the network error (and by how much) for a Single Layer Perceptron

Totalpoäng: 2

32 Q5 S4

Of the following, which is the **best** characterization of the Vanishing Gradient Problem
Choose one from below

- ○ Backpropagating error/loss does not work with functions for which derivative functions don't exist.

- ⦿ Backpropagating error/loss derivatives leads to increasingly small values for updating the weights the greater is the depth of the neural network. ✔

- ○ A problem for backpropagating error terms where the ANN is shallow

- ○ Backpropagating error/loss leads to increasingly small values for updating the weights the greater is the depth of the neural network.

Totalpoäng: 2

33 Q6 S4

Below is a confusion matrix which shows target class (ground truth) 1 = disease, 2 = no disease, and output class (ANN predictions) 1 = positive diagnosis, 2 = negative diagnosis.

What is the number of false negatives shown in the plot?

Choose one from below

○ 444

○ 234

◉ 14   ✅

○ 7

Totalpoäng: 2

34 Q7 S4

Below is a confusion matrix which shows target class (ground truth) 1 = disease, 2 = no disease, and output class (ANN predictions) 1 = positive diagnosis, 2 = negative diagnosis.

Given the ANN predicts the patient has the disease, what is the probability this is true?

Choose one from below

○ 0.944

◉ 0.984   ✅

○ 0.971

○ 0.969

Totalpoäng: 2

35 Q8 S4

For ANNs with hidden layers using backpropagation under what conditions will weights values stop updating:
Choose more than one from below

☑ When the loss on all data samples is zero ✅

☑ When the classification error on all data samples is zero. ❌

☐ When the weights gradients on all data samples are negative

☑ When the weights gradients on all data samples are zero ✅

Totalpoäng: 2

36 Q9 S4

Which of the below best describes the term "local minimum"?
Choose one from below

○ A point at which there is zero loss (over all data samples)

○ A point at which changes in the weights does not affect the loss or error of the network

○ A point at which changes in the weights does not affect the loss or error of the network ✔
   and the network loss is not at its potential minimum

◉ A point at which the network loss is not at its potential minimum ❌

Totalpoäng: 2

37 Q10 S4

When testing the overall quality of an ANN with a particular architecture, why is it important to carry out multiple runs with the same hyper-parameter settings?
Choose more than one from below (factors that can affect model performance)

☐ There may be different number of weights on each run

☑ Data is randomly partitioned into training, validation and test sets ✅

☑ Weights are randomly initialized ✅

☐ There may be a different number of hidden layers on each run

Totalpoäng: 2