Chair of Machine Learning of 3D Scene Geometry
School of CIT
Technical University of Munich

Esolution

Place student sticker here

Note:

- During the attendance check a sticker containing a unique code will be put on this exam.
- This code contains a unique number that associates this exam with your registration number.
- This number is printed both next to the code and to the signature field in the attendance check list.

Introduction to Deep Learning

Exam: IN2346 / Endterm **Date:** Friday 10th February, 2023
Examiner: Prof. Dr. Angela Dai **Time:** 18:30 – 20:00

	P 1	P 2	P 3	P 4	P 5	P 6	P 7	P 8
I								

Working instructions

- This exam consists of **20 pages** with a total of **8 problems**.
  Please make sure now that you received a complete copy of the exam.
- The total amount of achievable credits in this exam is 90 credits.
- Detaching pages from the exam is prohibited.
- **Answers are only accepted if the solution approach is documented.** Give a reason for each answer unless explicitly stated otherwise in the respective subproblem.
- Do not write with red or green colors nor use pencils.
- Physically turn off all electronic devices, put them into your bag and close the bag.

Left room from __________ to __________ / Early submission at __________

Problem 1 Multiple Choice (18 credits)

Mark correct answers with a cross ☒

To undo a cross, completely fill out the answer option ■

To re-mark an option, use a human-readable marking ×■

Please note:

- For all multiple choice questions any number of answers, i.e. either zero (!), one or multiple answers can be correct.
- **For each question, you'll receive 2 points if all boxes are answered correctly (i.e. correct answers are checked, wrong answers are not checked) and 0 otherwise.**

1.1 Your model for classifying different cat species is getting a low training set error with a high testing set error. Which of the following are promising things to try to improve your classifier?

□ Use a bigger neural network

□ Get more training data

□ Try a different initialization during training

□ Add weight regularization

1.2 Which of the following statements on activation functions are true?

□ The output values should be in the range of 0 to 1

□ Tanh can lead to vanishing gradients

□ Sigmoid outputs are zero-centered

□ Parametric ReLU can handle negative input values

1.3 Which of the following propositions are true about a Conv layer?

□ The total number of parameters depends on padding

□ The total number of parameters depends on the width and height of the input

□ The output depth is the same as the number of filters

□ The number of input channels and the number of filters' channels can differ

1.4 Logistic regression:

□ Allows performing binary classification.

□ Uses a variant of the cross entropy loss.

□ Can be seen as a 1-layer neural network.

□ The output space is between -1 and 1.

1.5 Regularization:

□ Is any technique that aims to reduce your validation error and increase your training accuracy.

□ Is any technique that aims to reduce the generalization gap.

□ Dropout, the use of ReLU activation functions, and early stopping can all be considered regularization techniques.

□ Weight decay (L^2) is commonly applied in neural networks to spread the decision power among as many neurons as possible.

1.6 What is the correct order of operations for an optimization with gradient descent?

(a) Update the network weights to minimize the loss.

(b) Calculate the difference between the predicted and target value.

(c) Iteratively repeat the procedure until convergence.

(d) Compute a forward pass.

(e) Initialize the neural network weights.

□ ebadc

□ bcdea

□ edbac

□ eadbc

1.7 So far we've learned Fully Connected Neural Network (FC), Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN). In which architecture does weight sharing occur across an input?

□ FC

□ CNN

□ RNN

□ None

1.8 Dropout...

□ ... makes your network train faster.

□ ... can be seen as an ensemble of networks.

□ ... is an efficient way for regularization.

□ ... has trouble with tanh activations.

1.9 Which of the following methods can be used in unsupervised learning?

□ Autoencoder.

□ PCA.

□ K-means.

□ Linear Regression.

Problem 2 Short Questions (19 credits)

0 □ □ 1 □ □ 2 □

2.1 Give one application scenario to use 1x1 convolution.

0 □ □ 1 □ □ 2 □

2.2 Explain the differences between binary classification and multiclass classification in terms of the output layer and loss function.

0 □ □ 1 □ □ 2 □

2.3 A convolutional neural network has 3 consecutive layers as follows:
5x5 Conv (stride 2) - 3x3 Conv (stride 2) - 3x3 Conv (stride 2).
How large is the receptive field of a pixel on the output? Note: Give it by MxM.

0 □ □ 1 □ □ 2 □

2.4 You are given a convolutional layer with kernel size 3, number of filters 3, stride 1 and padding 1. Compute the shape of the weights. Let's use the order of (Kernels, Channels, H, W) for the shape (0.5p). Write them down explicitly such that this convolutional layer represents the identity for an RGB image input. (1.5p).

0 □ □ 1 □ □ 2 □

2.5 Name one advantage and one disadvantage of Recurrent Neural Networks in general.

2.6 Briefly explain the concept of weight initialization of a neural network (1p). Name one bad method of initialization and explain why it is bad (1p). Additionally, name two common initialization strategies (0.5p each).

□ 0 □ □ 1 □ □ 2 □ □ 3

2.7 What is “early stopping”?

□ 0 □ □ 1 □ □ 2

2.8 Define “data augmentation" (0.5p), name two common data augmentation techniques used in image classification (0.5p each), and how could data augmentation be problematic in a supervised training scenario (1p)?

□ 0 □ □ 1 □ □ 2 □

2.9 Consider two different models for image classification of the MNIST data set.
The models are: (i) a 3 layer perceptron, (ii) LeNet.
Which of the two models is more robust to translation of the digits in the images? Give a short explanation why.

□ 0 □ □ 1 □

Problem 3 Backpropagation (8.5 credits)

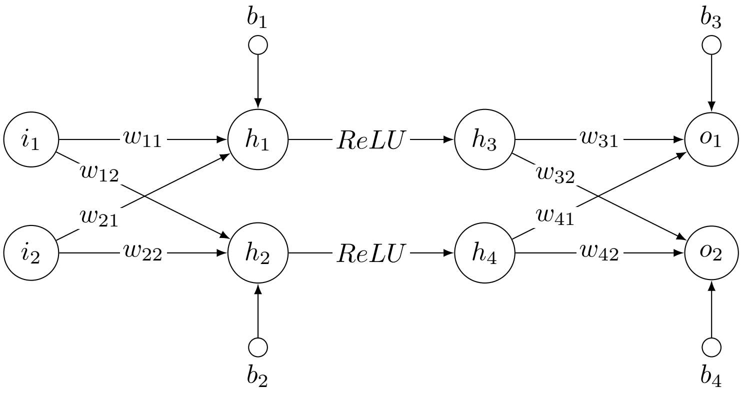


Figure 3.1: Simple network.

The values of variables are given in the following table:

Variable	i_1	i_2	w_{11}	w_{12}	w_{21}	w_{22}	w_{31}	w_{32}	w_{41}	w_{42}	b_1	b_2	b_3	b_4	t_1	t_2
Value	2.0	-1.0	1.0	-0.5	0.5	-1.0	0.5	-1.0	-0.5	1.0	0.5	-0.5	-1.0	0.5	1.0	0.5

0 □ □ 1 □ □ 2 □ □ 3 □

3.1 Compute the outputs (o_1 and o_2) of this network. Therefore, you will need to calculate the following variables: $h_1, h_2, h_3, h_4, o_1, o_2$.

0 □ □ 1 □ □

3.2 Write down the formula of the Mean Squared Error, and calculate the loss using your results in the previous question and the target values (t_1 and t_2). In case you have not solved the previous question, use the following values: $o_1 = 2$ and $o_2 = 0.5$.

3.3 Please update the weight w_{21} using gradient descent with learning rate $\alpha = 0.1$ as well as the loss computed previously. (Please write down all your computations.)

□ 0
□
□ 1
□
□ 2
□
□ 3
□
□ 4

Problem 4 Optimization (6 credits)

0 □
□
1 □

4.1 Explain the concept behind momentum in SGD.

0 □
□
1 □

4.2 Which optimizer introduced in the lecture uses second but not first order moment?

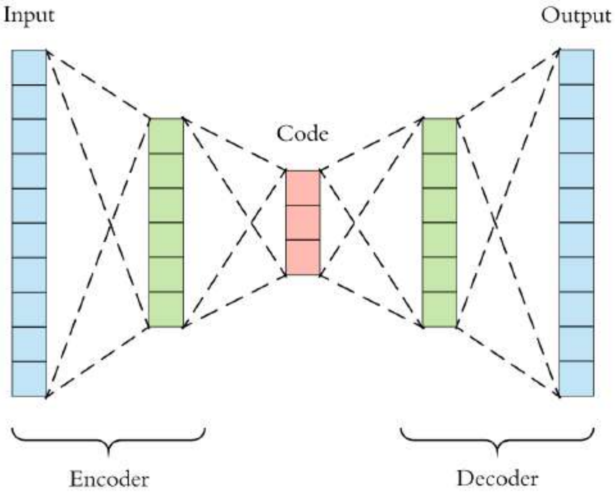
0 □
□
1 □
□
2 □

4.3 Name a disadvantage of a small minibatch/batch size and a disadvantage of a large minibatch/batch size.

0 □
□
1 □
□
2 □

4.4 Why is Newton’s method not commonly used in training a deep model (1p)? What would be an advantage of using it (1p)?

Problem 5 **Autoencoder (10 credits)**



5.1 How do each of the elements (encoder, code, decoder) of autoencoders function?

□ 0
□
□ 1
□
□ 2
□
□ 3

5.2 You want to perform a semantic segmentation task on a small labeled dataset, and you also have access to a larger unlabeled image dataset. Explain how an autoencoder can help in that given task.

□ 0
□
□ 1
□
□ 2

0 □ □ 1 □

5.3 If you use U-Net as your autoencoder model for semantic segmentation, what is a skip connection in the U-Net architecture? Don't forget to also offer some reasoning.

0 □ □ 1 □ □ 2 □

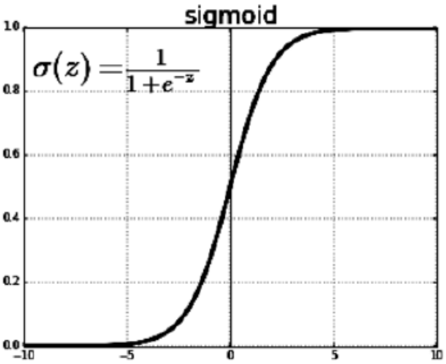
5.4 What are the differences between the autoencoder and the variational autoencoder in terms of the goal and loss?

0 □ □ 1 □ □ 2 □

5.5 The decoder part of an autoencoder can also be used in a Generative Adversarial Network (GAN). What is the difference between an autoencoder and a GAN in terms of network architecture? (0.5p each) What is the goal of using the discriminator loss in GAN? (1p)

Problem 6 CNNs (10 credits)

You are training a neural network with 10 convolutional layers with the non-linearity shown below:



6.1 Explain the behavior of the gradient of the non-linearity with respect to very large inputs.

□ 0 □ □ 1 □ □ 2

6.2 Why might this be a problem for training neural networks?

□ 0 □ □ 1 □ □ 2

6.3 Due to the problem mentioned in (6.2), modern architectures commonly adopt a different type of non-linearity. Name and draw this non-linearity, and explain why it helps solve the problem.

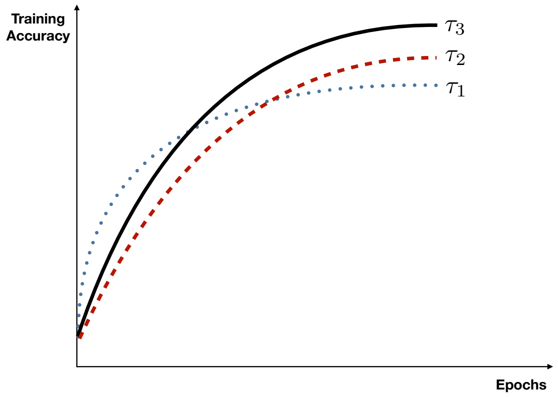
□ 0 □ □ 1 □ □ 2

6.4 You are training the network for image segmentation. After 50 epochs, you come to the conclusion that the network is too large for such a task. Name two approaches to counteract the problem, without changing the convolutional layers of your network.

□ 0 □ □ 1 □ □ 2

0 □
□
1 □
□
2 □

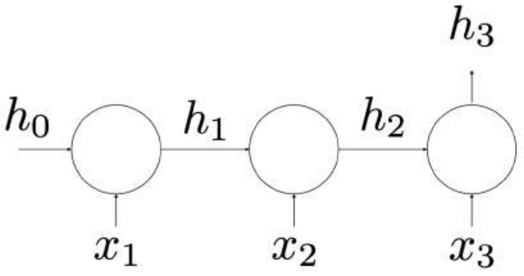
6.5 You adapt your network training accordingly, and now you are performing a grid search to find the optimal hyperparameters for vanilla stochastic gradient descent (SGD). You try three learning rates τ_i with $i \in \{1, 2, 3\}$, and obtain the following three curves for the training accuracy, all of the curves have already converged. Order the learning rates from larger to smaller.



Problem 7 LSTMs (9 credits)

7.1 Consider a vanilla RNN cell of the form $h_t = \tanh(V \cdot h_{t-1} + W \cdot x_t)$. The figure below shows the input sequence x_1, x_2, and x_3.

□ 0
□
□ 1
□
□ 2



Given the dimensions $x_t \in \mathbb{R}^4$ and $h_t \in \mathbb{R}^{12}$, what is the number of parameters in the RNN cell? Neglect the bias parameter.

7.2 If x_t is the 0 vector, then $h_t = h_{t-1}$. Discuss whether this statement is correct.

□ 0
□
□ 1
□
□ 2

7.3 Now consider the following **one-dimensional** ReLU-RNN cell.

□ 0
□
□ 1
□
□ 2
□
□ 3

$$h_t = \text{ReLU}(V \cdot h_{t-1} + W \cdot x_t)$$

(Hidden state, input, and weights are scalars)
Calculate h_1, h_2 and h_3 where $V = 1$, $W = 2$, $h_0 = -3$, $x_1 = 1$, $x_2 = 2$ and $x_3 = 0$.

0 □
□
1 □
□
2 □

7.4 A Long-Short Term Memory (LSTM) unit is defined as

$$\begin{aligned}
g_1 &= \sigma\left(W_1 \cdot x_t + U_1 \cdot h_{t-1}\right), \\
g_2 &= \sigma\left(W_2 \cdot x_t + U_2 \cdot h_{t-1}\right), \\
g_3 &= \sigma\left(W_3 \cdot x_t + U_3 \cdot h_{t-1}\right), \\
\tilde{c}_t &= \tanh\left(W_c \cdot x_t + u_c \cdot h_{t-1}\right), \\
c_t &= g_2 \circ c_{t-1} + g_3 \circ \tilde{c}_t, \\
h_t &= g_1 \circ c_t,
\end{aligned}$$

where g_1, g_2, and g_3 are the gates of the LSTM cell.
1) Assign these gates correctly to the **forget** f, **update** u, and **output** o gates. (1p)
2) What does the value c_t represent in a LSTM? (1p)

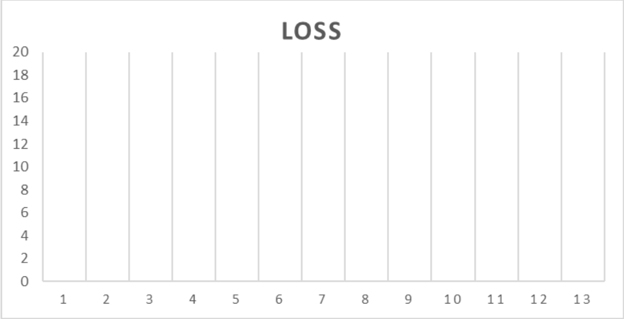
Problem 8 **Training & Evaluation (9.5 credits)**

8.1 A common way to divide your data is by splitting it into a train, validation, and test split. Explain the purpose of each split in detail and how we use each split (1p for each split). How much percentage of data do you commonly assign to each split (0.5p)?

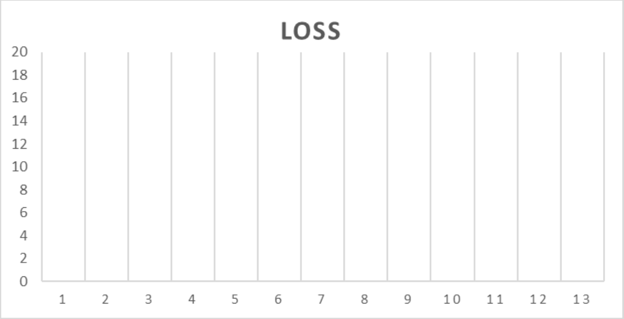
□ 0 □ □ 1 □ □ 2 □ □ 3 □

8.2 Explain the issues of overfitting and underfitting (1p each). Additionally, describe how your loss curves look like in each of the cases - draw the corresponding plots (1p each). (Make sure to label your curves).

□ 0 □ □ 1 □ □ 2 □ □ 3 □ □ 4



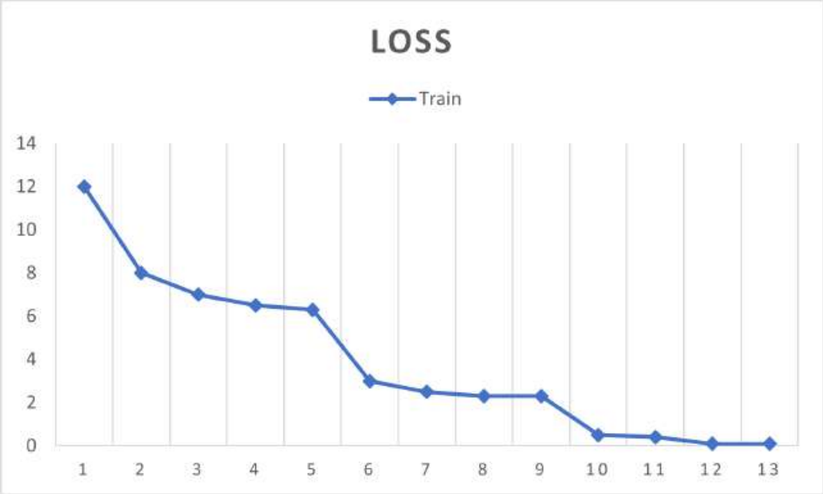
(a) Underfitting



(b) Overfitting

0 □ □ 1 □

8.3 A friend tries a new learning method and shows you this training loss plot. Name the method that was applied.



0 □ □ 1 □

8.4 You successfully trained your model on the task of Image Classification with product images you collected from Amazon. It achieves good classification accuracy on your collected data. Now, you took pictures of objects yourself, however, your model misclassifies most objects. Give one reason, why your model performs poorly on these images you took.

Additional space for solutions–clearly mark the (sub)problem your answers are related to and strike out invalid solutions.