Example exam questions

Deep learning for media technology, TNM112

1 Multiple choice questions

For each of the questions, there is only one alternative that is correct. You only need to check the correct alternative (no motivation needed).

1. What is the purpose of regularization:
   - ◯ Increasing the capacity of a model
   - ◯ Improving the generalization performance of a model
   - ◯ Making the optimization of a model faster
   - ◯ Reducing the complexity of a model
2. Suppose there are 50,000 training examples, and we train a neural network with a batch size of 200 samples for 3 epochs. How many times do we run the forward propagation?
   - ◯ 250
   - ◯ 750
   - ◯ 2,500
   - ◯ 7,500
3. If you notice a significant gap between the training accuracy and the test accuracy after training a neural network, which is a common method used to address this?
   - ◯ Dropout
   - ◯ Sigmoid activation
   - ◯ RMSprop optimization
   - ◯ Generative Adversarial Networks
4. Which of the following is not a Generative model?
   - ◯ Generative Adversarial Networks
   - ◯ Autoencoders
   - ◯ Variational Autoencoders
   - ◯ Latent Diffusion Models
5. Which of the following activation functions can lead to vanishing gradients?
   - ◯ Leaky ReLU
   - ◯ ReLU
   - ◯ Tanh
   - ◯ None of the above
6. What is the purpose of the softmax activation function:
   - ◯ It is used for regularization of a layer
   - ◯ It is used in SGD to enable the use of momentum
   - ◯ It is used to normalize the input datapoints provided to the network
   - ◯ It normalizes the final layer to provide a probability distribution function over class labels

2 Problems

For each of the problems, you need to clearly motivate your answer.

1. Given a dataset with **5×5 pixel** images (grayscale images), we want to perform classification into **2 different classes** (the output is 2 neurons with softmax activation). Calculate how many weights the following networks would contain:
   (a) An MLP with 3 dense hidden layers with 5 neurons each (i.e. 4 layers in total, 3 hidden + output layer).
   (b) A CNN with **1 convolutional layer**, with 5×5 filters and 3 output channels. Padding is applied to have the same resolution of output as input of the convolutional layer. The convolutional layer is followed by a **flattening layer**, and **a final dense layer** (mapping from flattened channels to class predictions).
2. We have a dense layer with 20 input activations and 10 output activations. How many parameters are added if we perform batch normalization on this layer? How many of these are trainable?
3. Formulate the following loss functions:
   (a) The L2 loss for an autoencoder (AE) with input x, encoder $E()$, and decoder $D()$.
   (b) The cross-entropy loss for a classifier network $f()$ with input x and one-hot encoded label y.
4. We have a training set with 5,000 datapoints. How many training steps are needed in the following cases:
   (a) We train for 10 epochs with a batch size of 50.
   (b) We train for 15 epochs with a batch size of 100.
5. Given the gradient of the loss function, $\nabla_W \mathcal{L}$, with respect to the network weights W, and learning rate α, formulate how an update of the weights can be performed in order to decrease the loss. How is this calculated for a batch of datapoints?
6. Compute the convolution (cross-correlation) operation given the following input and kernel:

Input array

4	-1	5	0	0
1	5	0	2	4
7	2	1	0	0
5	2	-1	3	0
-1	4	0	0	-4

$*$

Kernel

0	1	0
1	0	1
0	0	0

7. How many weights will the network specified in Keras below contain?

```
model.add(layers.InputLayer(input_shape=(6,6,3)))
model.add(layers.Conv2D(4, kernel_size=(3, 3), padding='same'))
model.add(layers.Conv2D(8, kernel_size=(3, 3), padding='same'))
model.add(layers.MaxPooling2D(pool_size=(2, 2)))
model.add(layers.Flatten())
model.add(layers.Dense(5, activation='softmax'))

```

8. Given dataset $\{x_i\}_{i=1}^n \subset \mathbb{R}^2$, $x_i = [x_1, x_2]^T$. Assuming we have a single layer MLP with weights W and biases b, where

$$W = \begin{bmatrix} W_1 & W_2 \\ W_3 & W_4 \end{bmatrix} \qquad b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix},$$

Given that the decision boundary is $x_1 = 2x_2 + 1$, find the possible values of weight and bias matrices that satisfy the decision boundary. Can you find more than one solution? If yes, can you give a more general solution?

3 Open answer questions

Questions can be answered in many different ways, and will be assessed based on how well you are able to demonstrate that you understand the different topics.

1. Explain in a few sentences what SGD is and how it is performed.

2. Explain in a few sentences how backpropagation works (equations not strictly necessary). What are the motivations and implications of backprop?

3. You have a dataset **D1** with 1 million labelled training examples for classification, and dataset **D2** with 100 labelled training examples. Your friend trains a model from scratch on dataset **D2**. You decide to train on **D1**, and then apply transfer learning to train on **D2**. State one problem your friend is likely to find with their approach. How does your approach address this problem?

4. Given a CNN classifier with convolutional layers followed by an MLP head, is it possible for this classifier to process arbitrary shapes of images? Motivate your answer.

5. How does the decision boundary look like for an MLP classifier without activation function?

6. Why are MLPs not the best option when it comes to image classification tasks?

7. Explain why Generative Adversarial Networks (GANs) are highly unstable to train.

8. How does splitting a dataset into train, validation, and test sets help identify overfitting?

9. Give a short overview of how generative diffusion models work. How can these be used to generate new data (e.g. images)?

10. Give three examples of regularization strategies. How and why can these improve generalization?

11. Give an example of a technique for explainable AI (XAI), with a high-level description of how it works. In what way can this technique help in understanding a neural network?

12. Give a short summary of how neural radiance fields (NeRFs) work. How is the training performed? What type of data is necessary to train the NeRF? How can the NeRF be used after training?

13. You have trained a neural network to classify Cat and Dog images using the CatsvsDogs dataset from Kaggle. You wanted to evaluate your model to see how it performs in unseen data. So, you used two test datasets. Dataset 1 is the test set from the CatsvsDogs data that wasn't used during training. Dataset 2 contains some images of Cats and Dogs captured using your mobile phone. After evaluation, you found out that your model achieves 95% accuracy on Dataset 1 and 60% accuracy on Dataset 2. What do you think might be the reason for the difference in performance in the two datasets?