DATA.ML.200 Deep Learning
Example electronic EXAM questions

These questions illustrate the types of questions that can be included in the electronic EXAM. Other types of questions (especially essay questions) will also be used. Exactly similar questions will be not used in the exam, so it is highly recommended to not memorize these answers.

1. Multiple choice questions

Loss Functions. Some of the statements below are TRUE and some of them are FALSE. Select all the TRUE statements.

- Cross-entropy loss is commonly used for classification tasks.
- Mean squared error is typically preferred over cross-entropy for multi-class classification.
- The use of cross-entropy loss leads to non-linear decision boundaries.
- L2 regularization penalizes large weights by adding their squared magnitude to the loss.

2. Multilayer perceptron

*The yellow boxes below indicate a set of possible values which would have given full points. In the loss calculation and the following steps, also 0.5*MSE was accepted. Partial points were given for partially correct answers (e.g., if the sign of the gradient was incorrect).*

Multilayer perceptron model

A multilayer perceptron consists of an input layer with two input notes, a hidden layer with three nodes, and one output node. The weight matrix of the first layer is
1 2
4 3

5 1
and the biases are 3, 2 and 1.
The weights of the second layer are 1, -2, and 3, and the bias is 4.
The nonlinearities of the hidden layer and output layer are rectified linear units (ReLU).

Forward pass

The MLP receives an input sample with feature values -1 and 2, for which the target output is 4. Compute the following quantities:
The pre-activations of the three hidden layer nodes are 6 , 4 , and -2 .
The activations of the three hidden layer nodes are 6 , 4 , and 0 .
The pre-activation of the output node is 2 .
The activation of the output node and the model output is 2 .

Loss calculation

The mean square error for the input sample is 4

Backward pass

The partial derivatives of the loss function with respect to the second layer weights are -24 and -16, and 0 with respect to the bias of the second layer is -4 .

Gradient descent

Do one gradient descent update step of the second layer weights and bias, with step size 0.1. The updated second layer weights are 3.4, -0.4, and 3 the updated bias is 4.4.

3. Convolutional Neural Network

The yellow boxes below indicate a set of possible values which would have given full points.

A convolutional neural network takes a grayscale image whose size is 100 x 100 pixels as an input. The size of convolutional kernels at the first convolutional layer is 3 x 3, and there are 50 kernels. The stride is 1 in both spatial dimensions, and no padding is

used. The spatial dimensions of the output feature map are 98 times 98, and the number of channels is 50 .
When bias terms are not counted, the number of learnable parameters in the convolutional layer is 450 .
The convolutional layer is followed by a max pooling layer that uses a 2x2 window and stride 2 in both spatial dimensions. The spatial dimensions of the resulting feature map are 49 times 49 , and the number of channels is 50 .
The number of learnable parameters in the max pooling layer is 0 .