183.663 Deep Learning for Visual Computing

Exam Questions and Information

Christopher Pramerdorfer

Computer Vision Lab, TU Wien

December 31, 2018

Every exam consists of a subset of the questions listed in this document. The exam questions will be very similar to those stated here, but not necessarily identical. You can answer the questions in English or German, but please use the English terminology in any case. Questions should be answered in your own words and in sufficient detail; do not just write down keywords or phrases from the lecture slides. You have 60 minutes to answer the questions. No documents are allowed during the exam.

What is the task definition of image classification? Explain at least 5 challenges and give examples. What is object detection and how does it differ from classification?

Why do we need datasets? Explain the purpose of the three different subsets covered.

Assume a company asks you to develop an application that is able to predict which kind of bird is depicted in a given image. Which kind of task is this? List and explain the individual steps you'd follow to solve this problem using deep learning.

What is the motivation for solving vision tasks via machine learning? What is a machine learning algorithm and how are they used for solving image classification problems?

What is a hyperparameter? Name at least 3 hyperparameters in the context of deep learning using convolutional neural networks. What is the purpose of hyperparameter selection, which search strategies exist, and how do they work?

How does the k nearest neighbor classifier work? Create a sketch for illustration, assuming a two-dimensional feature space and two different classes. Draw at least three

training samples per class (must not lie on a line) as well as (roughly) the resulting decision boundaries. What are the limitations of this classifier?

Why do general machine learning algorithms (those expecting vector input) perform poorly on images? What is a feature, and what is the purpose of feature extraction? Explain the terms low-level feature and high-level feature.

List and explain the steps of the traditional image classification pipeline. What are the differences to a corresponding deep learning pipeline?

What is the definition of a parametric model? What do the parameters of such models control (what effect do they have?), and how are they set?

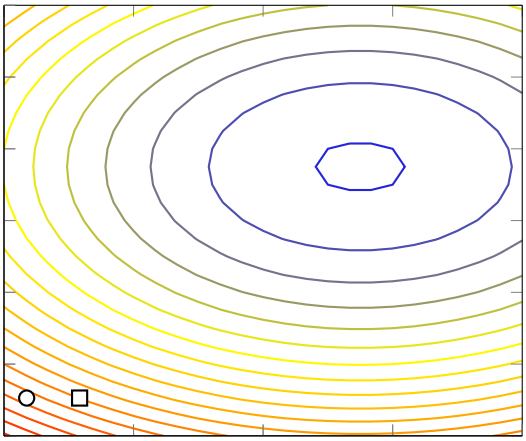
What is a linear model, which types of parameters does it have, and what do they specify? Draw a sketch assuming two-dimensional feature space and three different classes. Draw a few samples per class so that the classes are linearly separable. Draw the decision boundaries a linear classifier might learn and explain how the individual boundaries are related to the classifier output (no need to calculate anything).

What is the purpose of a loss function? What does the cross-entropy measure? Which criteria must the ground-truth labels and predicted class-scores fulfill to support the cross-entropy loss, and how is this ensured?

What is the purpose of optimization in the context of machine learning? How does the gradient descent algorithm work? What is the gradient of a function?

What is the difference between a local and a global optimum? Draw a sketch that illustrates the difference. Are local minima a problem in deep learning? Why (not)? What is momentum and why is it beneficial?

Consider the following contour plot of a function with two parameters. How might gradient descent proceed in this case, assuming the circle in the bottom-left corner as the starting point? Mark the individual steps and connect them using lines. Give a brief explanation of momentum. How would momentum affect the training progress in the example case? Mark the individual steps gradient descent with momentum might take, assuming the bottom-left square as the starting point?



Explain the differences between batch, minibatch, and stochastic gradient descent. Which version is most commonly used in deep learning and why? What effects has the minibatch size? Write pseudo-code that illustrates the overall structure of minibatch-based training and validation, continuing below the following line. A high-level overview is sufficient, no need to use math.

```
while epoch <= MAX_EPOCHS:
```

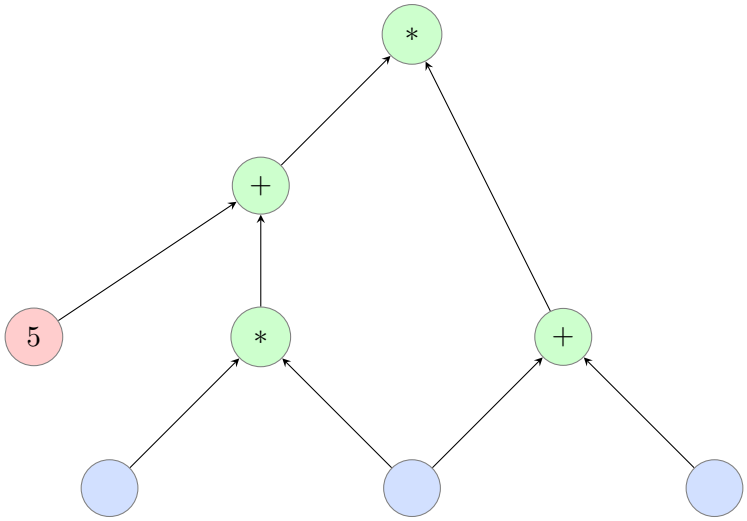
What is the definition of a feedforward neural network? Which types of units do such networks have? Draw a graph of such a network. How can linear models be implemented using neural networks?

What is the definition of a multilayer perceptron? What operation do (non-input) units perform? What is an activation function and which functions are common? Draw a sketch that shows the layers of such networks and how the individual units are connected. Why are multilayer perceptrons not suitable for deep learning for image analysis?

Explain the two ways covered for computing the gradient of loss functions as well as their pros and cons. Math is not required but of course allowed.

What is the purpose of the backpropagation algorithm, how does it differ from the "naive" algorithm for the same purpose, and what are its advantages? Explain the steps of the algorithm at a given node of a compuational graph.

Assume the following computational graph. First insert digits from your Matrikelnummer into the empty input nodes, going from right to left both in terms of nodes and digits. (Assuming Matrikelnummer 0123456, the values of the rightmost node would be 6, that of the node left of it would be 5, and so on.) Then compute the partial derivative of the topmost node with respect to all input nodes via backpropagation. Write computation node values after the forward pass left of the nodes, local gradients left of the edge connecting the corresponding nodes, and "cached" partial derivatives of the topmost node right to the nodes.



What is the motivation for and purpose of representation learning? How is deep learning related to representation learning, and what is its definition?

What is the receptive field of a neuron? Assume a network consisting of two convolutional layers with 3×3 connectivity followed by a 2×2 max-pooling with stride 2 and again two convolutional layers with 3×3 connectivity. What is the receptive field of neurons in the final convolutional layer? How does the receptive field affect feature extraction?

What is the purpose of convolutional layers and which operation do they compute? What are the two key differences to linear layers, and what motivates these differences with

respect to image analysis? What's the most popular activation function for these layers? Draw a graph of this function.

What are convolutional layers? Assuming an input shape of $W \times H \times D$, how many weight and bias parameters does a convolutional layer with a 3×3 kernel, F feature maps, stride 1, and padding have? What are feature maps and why are they needed?

What is the purpose of pooling layers? Calculate the output of a 2×2 max-pooling layer with stride 2 assuming the following input. What is global average-pooling and what is it used for? Are there alternatives to pooling layers?

$$\begin{bmatrix} 1 & 1 & 2 & 4 \\ 5 & 6 & 7 & 8 \\ 3 & 2 & 1 & 0 \\ 1 & 2 & 3 & 4 \end{bmatrix} \quad \Longrightarrow$$

Give a general overview of convolutional neural networks and their purpose, and draw a sketch that illustrates their overall structure (typical layer types and their arrangement). What are the two overall stages of such networks? What is needed to "combine" these stages / to make them compatible and how can this be achieved?

How is the depth of a CNN defined? What effect does increasing the depth have? Assuming an image classification problem with 10 classes and an image resolution of 64×64 pixels, specify a suitable CNN architecture using the notation Cx (3×3 convolution with x feature maps), Lx (linear layer with x neurons), R (ReLU), P (2×2 max-pooling), B (batch normalization). For instance "C16 R P L20" would mean "convolutional layer with 16 feature maps followed by ReLU, max-pooling and a linear layer with 20 neurons". Explain why you selected this architecture.

Why should images be normalized and how does this work during training and testing? What is the purpose of batch normalization? Between which layers/operations should batch normalization be applied?

What are the goals of optimization and machine learning? Why do they differ? Create two sketches with each showing the training progress over time in terms of both training and test error; one that is good from an optimization perspective but bad from a machine learning perspective, and one that is worse from an optimization perspective but better from a machine learning perspective. Explain both sketches.

What is early stopping and how does it work? What is the purpose of data augmentation? Assume that the task is to train a digit classifier. Think of and explain data

transformations that are applicable in this case, and at least one that is not.

What is the purpose of regularization? What is dropout, how does it work, and why is it effective? Where inside a network is dropout usually applied? What is weight decay and how does it differ from dropout?

What is transfer learning? How would you utilize transfer learning for solving an image classification problem via deep learning, assuming the available dataset is small. Are there issues with respect to data compatibility?

What is object detection? List two practical applications. What is the sliding window approach, how is it implemented in the context of deep learning, and which limitations does it have?

What is the basic idea of R-CNN and how does it differ from the sliding window approach? What is a region proposal? What is RoI Pooling and why is it needed?

What is Fast R-CNN, how does the overall pipeline look like, and how does it differ from R-CNN? How is the CNN trained and integrated?

What is Faster R-CNN, how does the overall pipeline look like, and how does it differ from Fast R-CNN? What is the purpose of region proposal networks, what do they predict, and how?

How does YOLO process images? Assuming two anchor boxes and five classes, what would YOLO predict? How is this information used for object detection / which post-processing steps are there?

What is semantic image segmentation? Draw a sketch of how CNNs for this purpose look like. What are the two overall stages of such networks? Which layers might they include that are not part of networks for other tasks?

What are transposed convolution layers? Assuming the following input I and kernel K, what would be the output of such a layer look like (stride 2, pad 1)?

$$I = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \quad , \quad K = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix} \qquad \Longrightarrow$$

Guest lecture questions will be added after these lectures.