UNIVERSITEIT TWENTE.

Mathematics of Imaging & AI
Department of Applied Mathematics, EEMCS
Prof. Dr. Christoph Brune
Dr. Gautam Pai, Mei Vaish

Exam: Deep Learning - From Theory to Practice (201800177)

Date: Thursday, Jan 22, 2026
Time: 13:45-15:45

- The exam has 28 points in total (grading scheme below).
- For exercise 1, write down your unique choice on your empty white paper (not this exercise sheet). For the other exercises, an explanation is required. Answers can be short and focused.
- No additional materials or tools are allowed. Good luck and success!

Exercise 1. (General concepts in Deep Learning) [6pt]

(a) **[1pt]** What does it mean that a network is robust against adversarial attacks?

(i) That the test loss is zero.
(ii) That random noise on the input won't change the output in most cases.
(iii) That the network generalises well to different datasets.
(iv) That any small distortion for a data-point does not significantly change the output.

(b) **[1pt]** Which of the following would you consider to be a valid activation function to train a nonlinear neural network using gradient descent?

(i) $f(x) = 0.1x + 1$
(ii) $f(x) = \max(x, 0.2x)$
(iii) $f(x) = \begin{cases} 0 & \mid x \geq 0 \\ 1 & \mid x < 0 \end{cases}$

(c) **[1pt]** For a standard 2D-convolutional layer, which property always holds?

(i) Translation equivariance
(ii) Translation invariance
(iii) Permutation equivariance
(iv) Permutation invariance

(d) **[1pt]** Which regularization method uses the prior that a neural network learns relevant features before fitting the noise?

(i) L^2-regularisation
(ii) Early stopping
(iii) Using an ensemble of models
(iv) Dropout

(e) **[1pt]** What is the main goal of an Auto-Encoder?

(i) To model sequential data.
(ii) To cluster high-dimensional data.
(iii) To learn features from local regions of the input.

(iv) To create art by generating images.

(f) **[1pt]** What problem could arise when using a Tanh activation function in a shallow neural network?

(i) Exploding gradients for small inputs

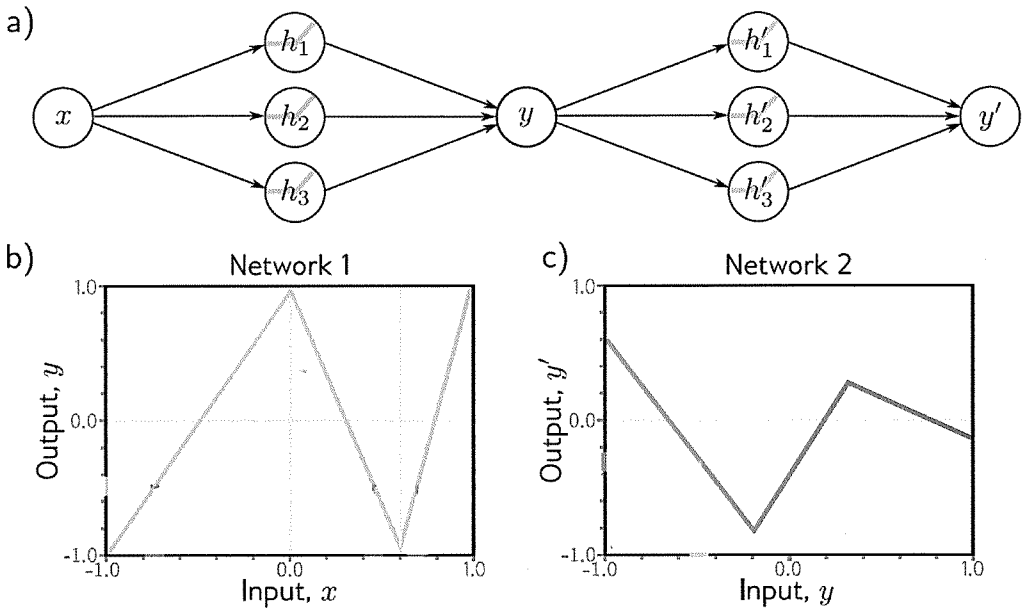
(ii) Exploding gradients for large inputs

(iii) Vanishing gradients for small inputs

(iv) Vanishing gradients for large inputs

Exercise 2. (Deep Neural Networks) [4pt]

(a) **[1.5pt]** Consider composing the two neural networks to get the network in figure (a). This first layer is plotted in figure (b), and the second layer is plotted in figure (c). Draw a plot of the relationship between the input x and output y' for $x \in [-1, 1]$.



(b) **[1.5pt]** Name 3 hyperparameters that are needed to define a convolutional or pooling layer.

(c) **[2pt]** Consider a network with $D_i = 5$ input neurons, $D_o = 5$ output neurons, and $K = 3$ layers, with $D = 20$ hidden units in each. The network contains no biases. By how much does the number of weights increase if we increase the depth by one? And by how much if we increase the width by one instead?

Exercise 3. (Neural Networks for Time Series) [8pt] We have time series data $x(t_1), \cdots, x(t_n) \in \mathbb{R}^d$ and want to learn a scalar feature $y \in \mathbb{R}$. There are many different architectures one could use.

One architecture is a recurrent neural network, with a hidden neuron $h \in \mathbb{R}^m$, given by

$$\begin{aligned} h_i &= \sigma(Wx(t_i) + Vh_{i-1}) \\ y &= Uh_i \end{aligned} \tag{1}$$

Here σ is some piecewise nonlinearity (for example ReLU).

(a) [**1pt**] Give the dimensions of the weight matrices W, V, U.

(b) [**2pt**] Write a recursive formula for the gradient of $\frac{dy}{dh_{i-1}}$, i.e. the right-hand side should contain the term $\frac{dy}{dh_i}$. Check that the dimensions match.

(c) [**1pt**] Suppose we train the network using weight decay, which makes the weights small. What can happen in this case to the gradient $\frac{dy}{dh_1}$ and why is it a problem?

(d) [**2pt**] We can change this recurrent network into a residual network, by adding skip connections. How does the formula (1) look like when we add skip connections? How does this change the gradient $\frac{dy}{dh_{i-1}}$ of question (b)?

(e) [**1pt**] Why does the problem mentioned in question (c) for RNNs, not occur in ResNets?

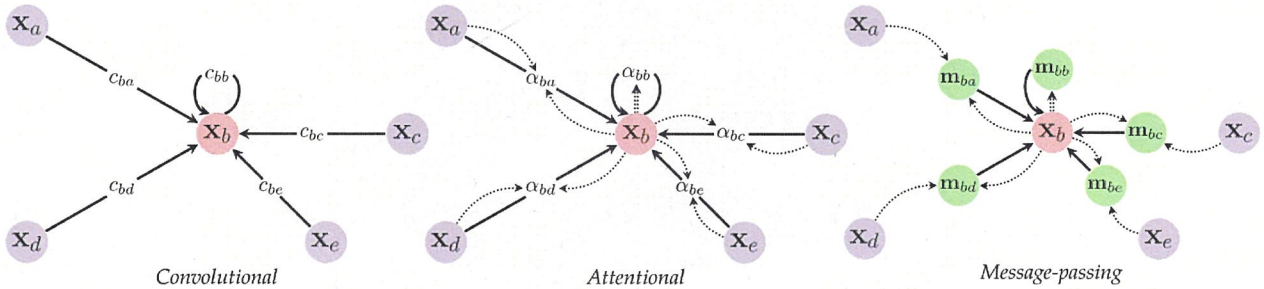
(f) [**1pt**] Suppose the data points $x(t)$ are sampled irregularly, i.e. $t_i - t_{i-1}$ is not constant. Why is this a problem for both the RNN and ResNet defined above? Name an architecture that is suitable for this type of data and explain why.

Exercise 4. (AutoEncoders, Generative Networks) [4pt]

(a) **[2pt]** Give a basic description of an autoencoder and explain why it can find a lower-dimensional representation.

(b) **[2pt]** What does a variational autoencoder (VAE) add to this architecture class as a key feature and how?

Exercise 5. (Graph Neural Networks) [5pt] There are three main 'flavours' to construct a graph neural network layer, each of which balances the need for expressiveness with a scalable implementation.



$$\mathbf{h}_i = \phi\left(\mathbf{x}_i, \bigoplus_{j \in \mathcal{N}_i} c_{ij}\psi(\mathbf{x}_j)\right) \qquad \mathbf{h}_i = \phi\left(\mathbf{x}_i, \bigoplus_{j \in \mathcal{N}_i} a(\mathbf{x}_i, \mathbf{x}_j)\psi(\mathbf{x}_j)\right) \qquad \mathbf{h}_i = \phi\left(\mathbf{x}_i, \bigoplus_{j \in \mathcal{N}_i} \psi(\mathbf{x}_i, \mathbf{x}_j)\right)$$

(a) **[2pt]** All three of these layers are permutation equivariant. What is permutation equivariance, and what is the benefit of this type of symmetry for learning on graphs?

(b) **[2pt]** Which layer type is associated with a transformer? Which layer type leads to the most scalable implementation of a graph neural network? Relate to the layer formulas $\mathbf{h}_i$ in your answers.

(c) **[1pt]** Which layer is the most expressive, i.e., it can represent the most functions.

Grading scheme:

Ex 1.	(a) 1pt	**Ex 2.**	(a) 1.5pt	**Ex 3.**	(a) 1pt	**Ex 4.**	(a) 2pt	**Ex 5.**	(a) 2pt
	(b) 1pt		(b) 1.5pt		(b) 2pt		(b) 2pt		(b) 2pt
	(c) 1pt		(c) 2pt		(c) 1pt				(c) 1pt
	(d) 1pt				(d) 2pt				
	(e) 1pt				(e) 1pt				
	(f) 1pt				(f) 1pt				

Total: 28 points

Disclaimer: Some questions in this exam have been partially generated by GPT, based on previous exam questions and human input. All questions have been edited, critically reviewed, and revised by the lecturers to make sure that they correspond to the learning goals of the course.