1 Feedforward NNs

Aufgabe 1: Stochastic Gradient Descent

1. What is the gradient of a neural network function f? (6 P)
2. Explain the method of gradient (or steepest) descent (initialization, update equation and parameter(s), conditions for convergence, type of solutions) (7 P)
3. Explain the method of stochastic gradient descent (especially the cost function and its properties, and mini-batches) (6 P)
4. Which problems does the method have? (6 P)

(25P)

Aufgabe 2: Feedforward Neural Networks

1. What is the objective function (the likelihood function) of a neural network for finding the k parameters of the network given m independent and identically distributed random variables? (6P)
2. Explain supervised and unsupervised learning algorithms and give for each type of algorithm the definition of the objective function which must be estimated by the neural network. (8P)
3. Explain the link between supervised and unsupervised learning algorithms! (4P)
4. Why do stochastic gradient descent algorithms scale for very large data sets? Explain the mini batch modification used in stochastic gradient descent algorithms! (7P)

(25P)

Aufgabe 3: Feedforward Neural Networks

1. Describe the architecture of a fully connected k-layer feedforward neural network! (8P)
2. Write down the matrix equations for computing the output of a two layer neural network! Do not forget to explain your notation! (8P)
3. Write down the pseudocode of forward propagation for a sparse neural network given as a computational graph! (9P)

(25P)

Aufgabe 4: Forward Propagation

1. Draw a fully connected neural network with 5 input nodes, 5 hidden nodes and 1 output node. Write down the equations for the ouputs of the input, hidden, and output layer! (4 P)
2. Write the pseudocode of the forward propagation algorithm. (6 P)
3. Which assumption does the forward propagation algorithm make on the order of nodes? (2 P)
4. Count the number of additions, multiplications and nonlinear function activations for the network you have drawn above for the forward propagation algorithm. (8 P)
5. What is the complexity of the forward propagation algorithm? (5 P)

(25P)

Aufgabe 5: Back-Propagation

1. What is the gradient of a function f?

   (5 P)
2. How is the Jacobian matrix of the function $f : R^m \to R^n : \mathbf{y} = f(\mathbf{x})$ with $\mathbf{x} \in R^m$, $\mathbf{y} \in R^n$ defined? (5 P)

3. A multilayer feedforward neural network can be symbolically as a path from input to output nodes: $a \rightarrow b \rightarrow c \rightarrow d \rightarrow e$ Or as a functional expression $g(f(f(f(a)$ with g the the function of the output layer and f the function of the hidden layers. What is the derivation of $\frac{\partial e}{\partial a}$? (10 P)
4. What is the problem with implementing this directly?

(25P)

Aufgabe 6: Feedforward Neural Networks

1. Describe the architecture of a fully connected k-layer feedforward neural network! (8P)
2. Write down the matrix equations for computing the output of a two layer neural network! Do not forget to explain your notation! (8P)
3. Write down the pseudocode of backward propagation for a fully connected k-layer feedforward neural network! (9P)

(25P)

Aufgabe 7: Classification Problems

1. Which activation function is usually used as output layer in a deep neural network which classifies a input data into k classes? Alternatives? (5 P)
2. Explain the formulas of the output layer of such a network! (5 P)
3. Which properties does this function have? (5 P)
4. Which problems does this function have? (10 P)

(25P)
Aufgabe 8: Hidden Layers

1. Which activation function is usually used for the hidden layers in a deep neural network which classifies a input data into k classes? Alternatives? (5 P)
2. Explain the formulas of one hidden layer of such a network! (5 P)
3. Which properties does this function have? (5 P)
4. Which problems does this function have? (5 P)
5. What is the derivative of the function? (And where is the derivative used?

(25P)
Aufgabe 9: Rectified Linear Units

1. How is the transformation function of a complete hidden layer with k hidden units defined? (4 P)
2. How is a rectified linear unit defined? How its gradient and its second derivative? (6 P)
3. What is the derivative of active units? What is the derivative of inactive units? (2 P)
4. Which problems do rectified linear units have? (5 P)
5. Give a geometric interpretation of the transformation by a rectified linear hidden layer! (8 P)

(25P)

Aufgabe 10: Regularization

1. Explain the purpose of regularization methods! (5 P)

2. Explain the basic idea of the dropout method! How is it integrated into the standard learning algorithm for feedworward neural networks? (5P)
3. Compare Dropout and Bagging! Which method do you prefer if your computational resources are relatively limited? (5P)
4. Describe and compare the L_2 and L_1 norm penalty methods! Which hyperparameter do these methods have? What are the effects of increasing the hyperparameter? How do you tune this hyperparameter? Which method should be used for model selection and why? (10P)

(25P)
Aufgabe 11: Regularization

1. What is Early Stopping and which problem of training machine learning algorithms does it solve? (8 P)
2. Write the pseudocode of an early-stopping algorithm! (8P)
3. Explain the basic ideas of three other regularization methods of your choice. (9P)

(25P)
Aufgabe 12: Regularization

1. Name and explain at least three methods of regularization based on the idea of perturbation. (6 P)
2. Explain the basic idea of the drop-out regularization method! (2 P)
3. Discuss parametrization and configurability of the drop-out method! (4 P)
4. What is a potential failure of the drop-out method? For which networks does the failure probably occur frequently? (2 P)
5. Explain Bagging (sampling method, training, decision rules). (6 P)
6. Why does Bagging work? (5 P)

(25P)
Aufgabe 13: Sentiment Analysis

1. What is sentiment Analysis? (Definition, explanation, and example). (6 P)
2. Draw a flow chart of the main steps of the data analysis process for sentiment analysis and explain each step! (8 P)
3. Explain the term frequency - inverted document frequency (TF-IDF) measure of a word for a document of a set of documents! (5 P)
4. How are documents preprocessed for neural networks? Explain the data set and how it is related to the neural network! What is a TF-IDF vector? (6 P)

(25P)

2 Genetic Algorithms

Aufgabe 14: Simple Genetic Algorithms

1. Draw the Flow diagram of a simple genetic algorithm! (5 P)
2. Describe the selection operator and draw a sketch of the fitness function if the probabilities of mutation and crossover are set to 0. (5 P)
3. Describe the crossover operator for fixed-length binary coded genes and draw a sketch of the fitness function if selection is elitist and uniform the probality of mutation is 0. Give a condition so that every possible bit-pattern can be represented when running this algorithm forever! (5 P)

4. Describe a mutation operator and its properties. (5 P)
5. How do you recognize that the optimum of the function you want to optimize cannot be reached? (5 P)

(25P)

Aufgabe 15: Simple Genetic Algorithms

1. What is a schema? How is its fitness defined? (5P)
2. What is the effect of reproduction on the number of representatives of a schema in the population (update formula and explanation)? (5P)
3. Explain the effect of crossover on a schema! What is the probability of survival of a schema for one-point crossover? (5P)
4. What are the main parameters of a genetic algorithm and how do you balance them properly? (5P)
5. How do you assess the quality of the solution of a genetic algorithm? (5P)

(25P)

Aufgabe 16: The Schema Theorem

1. What is a schema? How is its fitness defined? (5 P)
2. What is the effect of reproduction on the number of representatives of a schema in the population (update formula and explanation)? (5 P)
3. Explain the effect of crossover on a schema! What is the probability of survival of a schema for one-point crossover? (5 P)
4. What are the main parameters of a genetic algorithm and how do you balance them properly? (5 P)

(25P)

Aufgabe 17: Schema Theorem

We consider genetic algorithms with fixed-length binary coded genes.

- What is a schema? How is the value of a schema defined? Explain the main properties of a schema and their use in the analysis of the behavior of a genetic algorihtm! (5 P)
- How are successful schemata reproduced? (Formula and explanation!) (5 P)
- Explain the effect of crossover on the reproduction of genes! (5 P)
- Explain the effect of mutation on the reproduction of genes! (5 P)
- What is a deceptive problem? Explain! (5 P)

(25P)

Aufgabe 18: Implicit Parallelism and ...

1. What does *implicit parallelism* mean in the context of a genetic algorithm? (5 P)
2. And what are the effects of implicit parallelism? (5 P)
3. Visualize the 3-bit schemata $1, *, *$ and $*, *, 1$ for the function $y = x^2$ in the domain of $[-2, 2]$. (10 P)
4. How are schemata combined? (5 P)

(25P)

Aufgabe 19: Selection and Replication Dynamics

Given four individuals with a fitnesses of $f_1 = 10$, $f_2 = 20$, $f_3 = 22$, $f_4 = 100$ fill the probability of selection into the following table (8 P):

Selection Operator	I_1	I_2	I_3	I_4
Proportional to Fitness				
Proportional to Fitness Difference				
Linear Rank Selection				
Uniform Selection				

1. What is elitism? And what is the effect of elitism combined with a uniform selection operator on the convergence of the population of a genetic algorithm for a function maximization problem? (6P)
2. Explain the stochastic universal sampling (SUS) selection operator and its properties, especially with regard to bias and spread! (5P)
3. What is the problem of premature convergence? Compare selection operators of linear rank selection and selection proportional to fitness with regard to the problem of premature convergence! (6P)

(25P)

Aufgabe 20: Constraint Optimization

Let $0 \leq x_i \leq 10$ with $i \in 1, \ldots, n$ and let $n = 20$.
Minimize

$$f(\mathbf{x}) = -\left| \frac{\sum_{i=1}^{n} \cos(x_i)^4 - 2\prod_{i=1}^{n} \cos(x_i)^2}{\sqrt{\sum_{i=1}^{n} i x_i^2}} \right| \tag{1}$$

subject to:

$$c_1(\mathbf{x}) = -\prod_{i=1}^{n} +0.75 \leq 0 \tag{2}$$

$$c_2(\mathbf{x}) = \sum_{i=1}^{n} x_i - 7.5n \leq 0 \tag{3}$$

$$\tag{4}$$

1. How do you solve an optimization problem with constraints with a simple genetic algorithm? (8P)
2. What are the necessary conditions that $f(\mathbf{x}_*)$ is a minimum of the nonlinear programming problem? (6P)
3. Write the fitness function for the problem given above! (8P)
4. Is the solution of the genetic algorithm a global minimum? (3P)

(25P)

Aufgabe 21: Knapsack

You are the leader of a polar axpedition with 5 participants with 5 sleds and and 25 dogs. Your responsibility is the choice of food from the inventory of the base camp. All the food is packed in sealed containers which can not be opened and repacked. A list of all sealed containers with name of the food, weight of the container, and calories (in kcal) is available. The expedition is planned for a duration of 30 days. However, in the case of arctic storms, a longer duration is possible.
Every sled can take a maximum of 20kg of food. Every participant needs at least 2000 kcal per day, every dog needs at least 1500 kcal per day.

- Formulate the problem as a Knapsack problem. (6 P)
- You plan to solve the problem with a simple genetic algorithm:
  - What is the fitness function? (2 P)
  - Which types of constraints does the Knapsack problem have? (2 P)
  - How are the constraints handled in the genetic algorithm? (6 P)

- Unfortunately, your PC fails. How could you approximate the solution manually? (2 P)
- Under which conditions should you cancel the expedition? (2 P)
- What is the method of Lagrange? Explain the method! (5 P)

(25P)

Aufgabe 22: Real Representation

1. Which properties does an evolutionary algorithm need to ensure convergence to a globally optimal solution as the number of trials goes to infinity? (5 P)
2. Write the pseudocode of the classic differential evolution algorithm! (10 P)
3. Explain the mutation operation of a classic differential evolution algorithm! (10 P)

(25P)

Aufgabe 23: Permutation Representation

1. For which industrial problems is a gene representation in the form of a permutation of integers appropriate? Name at least three! Explain the generic TSP problem behind them. (5 P)
2. Describe a mutation operator for the permutation representation. (5 P)
3. Describe a crossover operator for the permutation representation. (5 P)
4. What is a hybrid genetic algorithm? (5 P)
5. Explain a suitable heuristic for a hybrid genetic algorithm for the TSP! (5 P)

(25P)

Aufgabe 24: TSP

M is an $n \times n$ asymmetric matrix whose off-diagonal elements contain distances (in time) between n cities. The goal is to find the shortest round-trip in which the traveller visits each city exactly once.

1. What is the fitness function of a genetic algorithm for this problem? (5 P)
2. What is a suitable gene representation for this problem? (5 P)
3. Define a mutation operator for the chosen gene representation! (5 P)
4. Define a crossover operator for the chosen gene representation! (5 P)
5. Explain the Lin-Kernighan heuristic! (5 P)

(25P)

Aufgabe 25: Meta Genetic Algorithms

1. Draw the Flow diagram of a meta genetic algorithm! (10 P)
2. Prepare a computational experiment with a meta genetic algorithm! You want to find out good parameter settings for a simple genetic algorithm with binary genes for optimizing the weights of feedforward neural networks using the forward propagation algorithm for k-bit Boolean functions for a reasonable range of k.
   (a) Which parameters of the genetic algorithm should you choose for your experiment? Which parameter ranges? (8 P)
   (b) Which tasks should you select? And why? (4 P)
   (c) How many trials and why? (3 P)

(25P)