

Deep Learning

Name:

Vorname:

Matr.-Nr.:

Prüfungsordnung:

Mit Ihrer Unterschrift bestätigen Sie, dass Sie gesundheitlich dazu in der Lage sind an der Klausur teilzunehmen.

Datum & Unterschrift:

Erlaubte Hilfsmittel: Taschenrechner

Mit **Bleistift** oder **in rot** geschriebene Aufgaben werden nicht gewertet. Die Bearbeitungszeit für die Klausur beträgt 90 Minuten. Schreiben Sie bitte auf jedes Blatt Ihren **Namen** und Ihre **Matrikelnummer** in die dafür vorgesehene Zeile.

Maximal erreichbare Punktzahl: 90.

Aufgabe	1	2	3	4	5	6	7	8	9	10	11	12	Σ
Erreichte Punkte													
Erreichbare Punkte	8	7	8	9	9	7	8	8	7	7	8	8	90

Aufgabe 1 – Lineare Algebra – 8 Punkte

(8 Punkte)

Gegeben seien die Matrizen:

$$\mathbf{A} = \begin{pmatrix} 2 & 1 \\ 0 & 3 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 1 & -1 \\ 2 & 0 \end{pmatrix} \quad (1)$$

- a) Berechnen Sie $\mathbf{A} \cdot \mathbf{B}$.

- b) Berechnen Sie \mathbf{A}^T .

- c) Beschreiben Sie, wie eine 2×2 Gewichtsmatrix zwei Eingaben mit zwei Neuronen verbindet.

Aufgabe 2 – Aktivierungsfunktionen – 7 Punkte

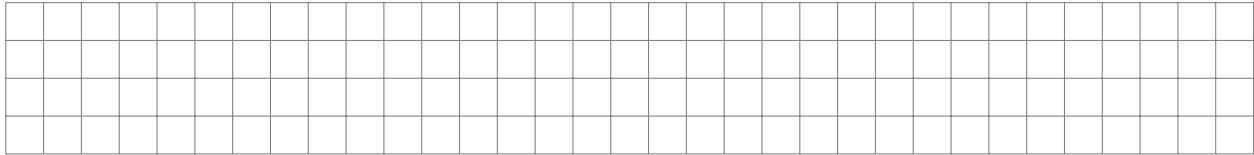
- a) Berechnen Sie die Ableitung der Sigmoid-Funktion $\sigma(x) = \frac{1}{1+e^{-x}}$.

- b) Berechnen Sie $\sigma(0)$.

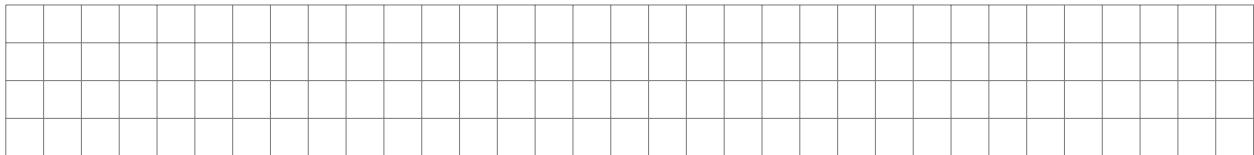
- c) Geben Sie die Ableitung der ReLU-Funktion $\text{ReLU}(x) = \max(0, x)$ an.

Aufgabe 3 – Das Perceptron und das XOR-Problem – 8 Punkte

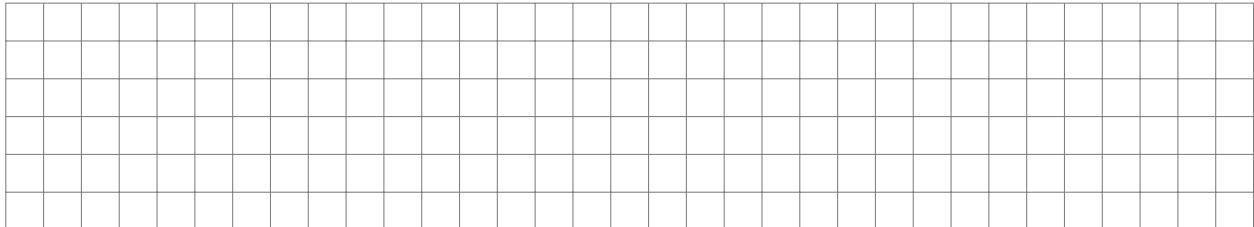
- a) Zeichnen Sie die Architektur eines Single-Layer-Perceptrons mit 2 Eingängen und 1 Ausgang.



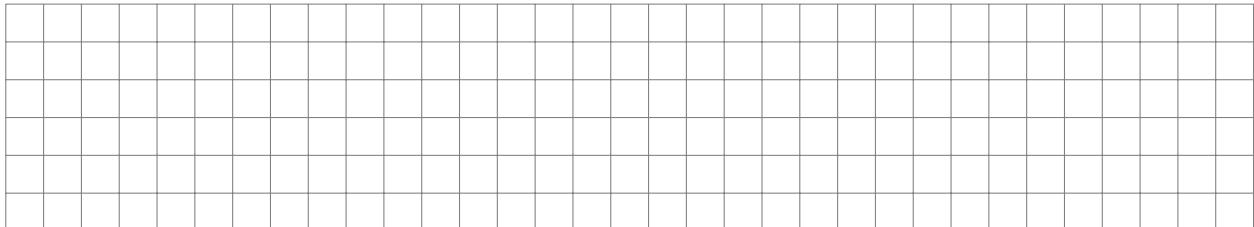
- b) Warum kann ein Single-Layer-Perceptron die XOR-Funktion nicht lösen?



- c) Zeichnen Sie ein Multi-Layer-Perceptron mit 2 Eingängen, 1 Hidden Layer mit 2 Neuronen und 1 Ausgang. Wie viele Gewichte?



- d) Warum kann dieses Netzwerk XOR lösen?



Aufgabe 4 – Partielle Ableitungen und Kettenregel – 9 Punkte

Gegeben: $L = \frac{1}{2}(y - a)^2$ mit $y = 1$, $a = \sigma(z)$, $z = w \cdot x + b$.

a) Berechnen Sie $\frac{\partial L}{\partial a}$.

b) Berechnen Sie $\frac{\partial a}{\partial z}$ für Sigmoid $\sigma(z)$.

c) Berechnen Sie $\frac{\partial z}{\partial w}$.

d) Nutzen Sie die Kettenregel: $\frac{\partial L}{\partial w} = \frac{\partial L}{\partial a} \cdot \frac{\partial a}{\partial z} \cdot \frac{\partial z}{\partial w}$.

e) Für $x = 1, w = 0.5, b = 0$: Berechnen Sie z , dann $\frac{\partial L}{\partial w}$ (mit $\sigma(0.5) \approx 0.62$).

Aufgabe 5 – Forward Pass eines 2-Schicht-Netzwerks – 9 Punkte

Gegeben: $\mathbf{x} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $\mathbf{W}^H = \begin{pmatrix} 0.5 & 0.3 \\ 0.2 & 0.4 \end{pmatrix}$, $\mathbf{b}^H = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

Output Layer: $\mathbf{w}^O = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $b^O = 0$

Hidden Layer: ReLU, Output Layer: Linear (keine Aktivierung).

a) Berechnen Sie $\mathbf{z}^H = \mathbf{W}^H \mathbf{x} + \mathbf{b}^H$.

b) Berechnen Sie $\mathbf{a}^H = \text{ReLU}(\mathbf{z}^H)$.

c) Berechnen Sie $\hat{y} = \mathbf{w}^O \cdot \mathbf{a}^H + b^O$.

d) Mit Label $y = 2$: Berechnen Sie den Verlust $L = \frac{1}{2}(y - \hat{y})^2$.

e) Berechnen Sie $\frac{\partial L}{\partial \hat{y}}$.

f) Berechnen Sie $\frac{\partial L}{\partial w_1^O}$ mittels Kettenregel.

Aufgabe 6 – Gradient Descent – 7 Punkte

- a) Schreiben Sie die Gewichtsaktualisierungsformel für Gradient Descent auf.

- b) Ein Gewicht $w = 0.5$ hat Gradienten $\frac{\partial L}{\partial w} = 0.4$. Aktualisieren Sie mit $\eta = 0.1$.

- c) Nennen Sie zwei Probleme bei zu hoher Learning Rate.

- d) Was ist der Unterschied zwischen Batch GD und SGD?

Aufgabe 7 – Convolutional Neural Networks (CNNs) – 8 Punkte

- a) Nennen Sie zwei Gründe, warum CNNs besser für Bilder sind als vollverbundene Netze.

- b) Was berechnet die Convolution-Operation?

- c) Zeichnen Sie ein Max-Pooling Beispiel: 2×2 Filter über einer 4×4 Feature Map.

- d) Nennen Sie zwei CNN-Architekturen und je eine Besonderheit.

Aufgabe 8 – RNNs und LSTMs – 8 Punkte

- a) Nennen Sie zwei Gründe, warum RNNs für Sequenzen besser sind als vollverbundene Netze.

- b) Was ist das Vanishing Gradient Problem bei RNNs?

- c) Skizzieren Sie eine LSTM-Zelle mit vier Komponenten.

- d) Warum lösen LSTMs das Vanishing Gradient Problem besser?

Aufgabe 9 – Generative Modelle – 7 Punkte

- a) Unterschied zwischen diskriminativen und generativen Modellen?

- b) Skizzieren Sie die Architektur eines GAN (Generator und Discriminator).

- c) Nennen Sie zwei Anwendungen von GANs.

Aufgabe 10 – Overfitting und Regularisierung – 7 Punkte

- a) Unterschied zwischen Underfitting, Good Fit und Overfitting (mit Skizze)?

- b) Nennen Sie drei Techniken zur Vermeidung von Overfitting.

- c) Was ist Dropout und wie funktioniert es beim Training?

- d) Wie ändert sich Dropout bei der Inferenz?

Aufgabe 11 – Gradienten in Vektorform – 8 Punkte

Gegeben: $L = \frac{1}{2}(y - \hat{y})^2$ mit $\hat{y} = \mathbf{w}^T \mathbf{a} + b$, $\mathbf{a} = \text{ReLU}(\mathbf{Wx} + \mathbf{b})$

- a) Berechnen Sie $\frac{\partial L}{\partial \hat{y}}$.

- b) Berechnen Sie $\frac{\partial \hat{y}}{\partial \mathbf{w}}$ (als Vektor).

- c) Nutzen Sie die Kettenregel: $\frac{\partial L}{\partial \mathbf{w}} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial \mathbf{w}}$.

- d) Gegeben $\mathbf{a} = \begin{pmatrix} 1 \\ 0.5 \end{pmatrix}$, $y = 2$, $\hat{y} = 0.75$: Berechnen Sie $\frac{\partial L}{\partial \mathbf{w}}$ numerisch.

- e) Warum normalisiert man den Gradienten bei Gradient Descent in die entgegengesetzte Richtung?

Aufgabe 12 – Momentum und adaptive Learning Rates – 8 Punkte

- a) Erklären Sie die Momentum-Methode kurz.

- b) Was ist der Unterschied zwischen Momentum und SGD?

- c) Was ist ADAM und welche Parameter hat es?

- d) Wann ist ADAM besser als Standard Gradient Descent?

- e) Was ist der Vorteil von adaptiven Learning Rates?