

Ausgewählte Kapitel sozialer Webtechnologien

Test-Klausur SS 2022

Name: _____

Matrikelnummer: _____

- Schreiben Sie auf jedes Blatt, das Sie abgeben, Name und Matrikelnummer
- Schreiben Sie sauber, nicht leserliche Abgaben können zu Punktabzug führen

Viel Erfolg!

Aufgabe:	1	2	3	4	5	6	Summe:
Punkte:	12	24	12	12	24	16	100
Ergebnis:							

1. (12 Punkte) Multiple-Choice-Fragen

Für jede der folgenden Fragen kreisen Sie bitte die Antwort(en) ein. Es können mehrere Antworten korrekt sein. Falsche Angaben bewirken Punktabzug (keine Negativpunkte möglich).

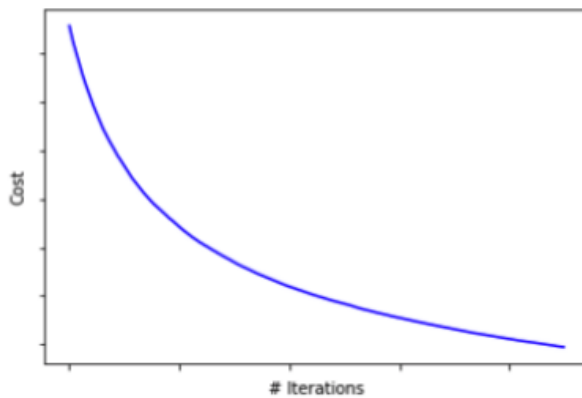
- (a) (2 Punkte) Um Backpropagation durch einen Max-Pooling-Layer zu machen, müssen Sie die Indizes der maximalen Werte aus dem Forward-Path wissen?
- (A) Wahr
(B) Falsch
- (b) (2 Punkte) Sie implementieren einen Klassifikator zum Erkennen von Krebs. Das Modell unterscheidet zwischen Tumor- (label 1) und Normal-Gewebe (label 0), es soll von Onkologen_innen am Krankenhaus verwendet werden. Welche der beiden Metriken würden Sie in diesem Kontext verwenden?
- (A) $\text{Precision} = \frac{\text{TruePositives}}{\text{TruePositives} + \text{FalsePositives}}$
(B) $\text{Recall} = \frac{\text{TruePositives}}{\text{TruePositives} + \text{FalseNegatives}}$
- (c) (2 Punkte) Um eine gute Generalisierung zu erreichen, sollte ein neuronales Netzwerk mindestens doppelt so viele lernbare Parameter besitzen, wie Sie Trainingsdaten zur Verfügung haben.
- (A) Wahr
(B) Falsch
- (d) (2 Punkte) Der Backpropagation-Algorithmus aktualisiert alle lernbaren Parameter in einem Netzwerk.
- (A) Wahr
(B) Falsch
- (e) (2 Punkte) Sie haben ein simples neuronales Netzwerk mit einem Convolutional-Layer. Welche der folgenden Eigenschaften treffen auf das Netzwerk zu?
- (A) Es ist Rotationsinvariant
(B) Es ist Translationsinvariant
(C) Es ist Skaleninvariant
(D) Alle Eigenschaften treffen zu
- (f) (2 Punkte) Welche Aussagen über Convolutional-Neural-Networks (CNNs) im Kontext der Bildanalyse sind korrekt?
- (A) Filter in den ersten Faltungsschichten entsprechen Kantendetektoren
(B) CNNs haben bei gleicher Anzahl von Neuronen und Schichten mehr lernbare Parameter als vollständig-vernetzte Netzwerke (fully-connected networks)
(C) Pooling-Schichten reduzieren die Auflösung eines Bildes (down sampling)
(D) CNNs eignen sich nicht für Regressionsprobleme

2. (24 Punkte) Verständnisfragen (kurze Antworten)

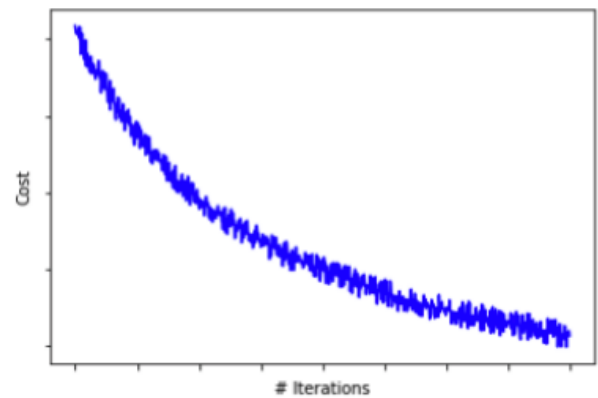
Geben Sie eine prägnante Antwort auf jede Frage in zwei bis drei Sätzen.

- (a) (3 Punkte) Sie haben als Eingabe ein Bild mit den Dimensionen $100 \times 100 \times 3$. Sie transformieren es in einen Vektor (image flattening) und nutzen es in einer vollständig vernetzten Schicht (fully-connected layer) mit 50 Neuronen. Welche Dimensionen haben die Gewichtsmatrix und der Bias-Vektor?
- (b) (3 Punkte) Sie senden ein Bild X durch ein neuronales Netzwerk. Die Ausgabe ist eine Wahrscheinlichkeit \hat{y} . Erläutern Sie was $\frac{\partial \hat{y}}{\partial x}$ bedeutet?
- (c) (3 Punkte) Wenn Ihre Eingabe aus nur zwei Dimensionen besteht, können Sie die Entscheidungsgrenze im euklidischen Raum darstellen und erkennen, ob Ihr Model über angepasst (overfitting) ist. Wie stellen Sie dies bei höher dimensionalen Eingaben, beispielsweise einem Bild, fest?
- (d) (3 Punkte) Angenommen, es folgen in einem Convolutional-Neural-Network auf eine Faltungsschicht $CONV1$ direkt drei weitere Faltungsschichten $CONV2$, $CONV3$ und $CONV4$. Alle Schichten haben eine Filtergrösse von $(3,3)$ und eine Stride $S = 1$. Welches effektive Receptive Field wird durch diese Stapelung auf dem Input von $CONV1$ realisiert?
- (e) (3 Punkte) Erläutern Sie weshalb Dropout eine Regularisierungstechnik ist?

- (f) (3 Punkte) Welche Probleme erwarten Sie, wenn eine zu hohe Lernrate verwendet wird? Wie können Sie dieses Problem feststellen?
- (g) (3 Punkte) Was ist ein Sattelpunkt? Welche Vor-/Nachteile hat das Mini-Batch Gradientenabstiegsverfahren (mini batch gradient descent) bei der Problematik von Sattelpunkten?
- (h) (3 Punkte) Die folgenden Grafiken zeigen, wie sich die Kosten für zwei unterschiedliche Optimierungsalgorithmen über mehrere Iterationen verringern. Erläutern und begründen Sie anhand der Kurvenverläufe welcher Algorithmus das Batch-Gradientenabstiegsverfahren (batch gradient descent) und welcher das Mini-Batch-Gradientenabstiegsverfahren (mini batch gradient descent) nutzt.



(a) Graph A



(b) Graph B

3. (12 Punkte) Lineare Algebra

Ihnen sind folgende Informationen zu einem neuronalen Netzwerk gegeben. Gewichtsmatrizen der Schichten, wobei das erste Element jeder Zeile einem Bias entspricht:

$$W_{Hidden} = \begin{pmatrix} 10 & 20 & -20 \\ 20 & 40 & 10 \\ 10 & -20 & -20 \end{pmatrix}$$

$$W_{Output} = (10 \quad 30 \quad -10 \quad -20)$$

Aktivierungsfunktion $g(z)$, die bei allen Neuronen gleich ist:

$$g(z) = \begin{cases} 1 & z \leq -10 \\ 0 & z \geq 10 \\ 0.5 & \text{sonst} \end{cases}$$

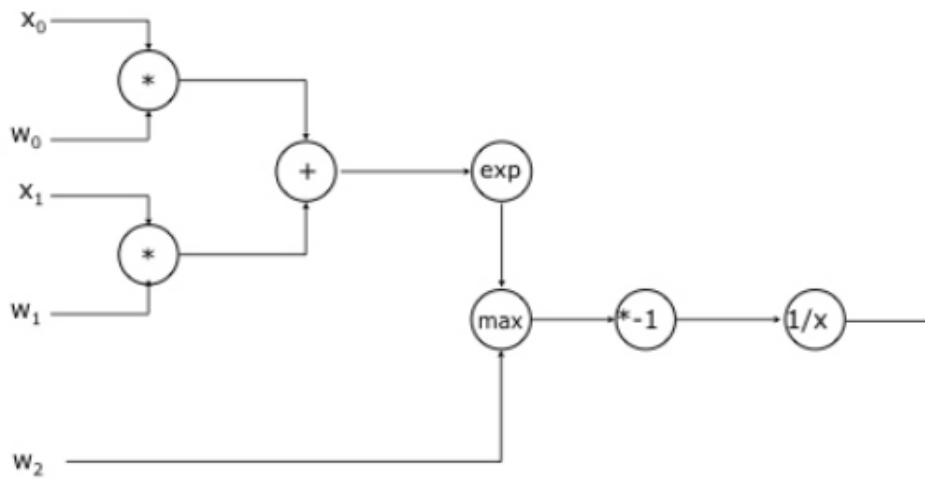
- (a) (4 Punkte) Zeichnen Sie das beschriebene neuronale Netzwerk mit allen Neuronen und deren Verbindungen auf. Notieren Sie Gewichte und Bias an die jeweiligen Verbindungen. Zeichnen Sie sauber, nicht leserliche Skizzen werden nicht gewertet.

- (b) (8 Punkte) Erstellen Sie aus den gegebenen Vektoren x_1, x_2, x_3 und x_4 eine Mini-Batch-Matrix. Berechnen Sie basierend auf diesem Input den Output des Netzwerks. Notieren Sie auf dem Weg ebenfalls die Präaktivitäten der verdeckten Neuronen, deren Aktivierungen und die Präaktivitäten des Ausgangs. Um die vollständige Punktzahl zu erhalten, müssen alle Schritte mittels Matrix-Operationen berechnet werden.

$$\vec{x}_{(1)} = [0, 0], \quad \vec{x}_{(2)} = [0, 1], \quad \vec{x}_{(3)} = [1, 0], \quad \vec{x}_{(4)} = [1, 1]$$

4. (12 Punkte) Backpropagation

Berechnen Sie die partiellen Ableitungen $\frac{\partial out}{\partial w_0}$, $\frac{\partial out}{\partial w_1}$ und $\frac{\partial out}{\partial w_2}$ mittels Backpropagation in folgendem gegebenen Computational Graph:



Dabei haben die Variablen folgende Werte:

- $x_0 = 4$ $w_0 = 1$
- $x_1 = -2$ $w_1 = 2$
- $w_2 = 0.5$

Hinweis: Exponentialfunktion: $\exp = e^x$ und die Ableitung $\frac{d}{dx}(e^x) = e^x$

5. (24 Punkte) Convolution

Gegeben sind exemplarisch die Werte von drei Kanälen in der linken unteren Ecke einer Aktivierungsschicht A^l . Die vollständigen Dimensionen des Volumens A^l beträgt $16 \times 16 \times 32$. Im Netzwerk wird ReLU als Aktivierungsfunktion verwendet und nach jeder Faltung durchgeführt.

$$A_1^l = \begin{pmatrix} \dots & \dots & \dots \\ 9 & 9 & \dots \\ 2 & 8 & \dots \\ 1 & 3 & \dots \end{pmatrix}, \quad A_2^l = \begin{pmatrix} \dots & \dots & \dots \\ 9 & 9 & \dots \\ 3 & 8 & \dots \\ 2 & 5 & \dots \end{pmatrix}, \quad A_3^l = \begin{pmatrix} \dots & \dots & \dots \\ 6 & 6 & \dots \\ 3 & 8 & \dots \\ 1 & 5 & \dots \end{pmatrix}$$

Folgende Werte haben die Gewichte der Kernel K_{ij} (Filter) und die Bias b_i des folgenden Faltungsschichtneurons $CONV^{l+1}$ (*convolutionallayer*).

- i : Index für den Ausgabekanal (*output featuremap*)
- j : Index für den Eingabekanal (hier A_1^l, A_2^l, A_3^l)

$$K_{11} = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}, \quad K_{12} = \begin{pmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{pmatrix}, \quad K_{13} = \begin{pmatrix} -1 & 0 & 2 \\ -2 & 0 & 1 \\ -1 & 0 & 1 \end{pmatrix}, \quad b_1 = 1$$

$$K_{21} = \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix}, \quad K_{22} = \begin{pmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}, \quad K_{23} = \begin{pmatrix} -2 & -2 & -2 \\ 0 & 0 & 0 \\ 2 & 2 & 2 \end{pmatrix}, \quad b_2 = -1$$

Die Kernel werden mit einer Stride $S = 1$ verschoben. Ferner wird Zero-Padding zur Erhaltung der räumlichen Struktur verwendet. Die Dimension in Breite und Höhe von A^l entspricht also der von A^{l+1} .

(a) (2 Punkte) Welchen Wert hat das Zero-Padding P um die räumlichen Struktur zu erhalten?

(b) (4 Punkte) In der Faltungsschicht $CONV^{l+1}$ befinden sich insgesamt 8 Kernel ($i \in 1, \dots, 8$). Welche Dimension hat das Ausgabevolumen A^{l+1} und wieviele lernbare Parameter enthält die Schicht $CONV^{l+1}$?

- (c) (18 Punkte) Berechnen Sie konventionell (nicht vektorisiert) alle Werte von A^{l+1} des Neurons $CONV^{l+1}$, die mit den gegebenen Informationen berechenbar sind. Achten Sie auf eine klare Darstellung der Vorgehensweise und des Rechenweges. Nicht nachvollziehbare Antworten können nicht gewertet werden.

6. (16 Punkte) Optimizer

- (a) (12 Punkte) Erläutern Sie das Gradientenabstiegsverfahren mit Momentum. Geben Sie eine anschauliche Beschreibung des Verfahrens an. Nennen Sie die Vorteile und Eigenschaften, welche das Verfahrens gegenüber dem Gradientenabstiegsverfahren ohne Momentum hat. Welches typische Verhalten zeigt das Verfahren während des Optimierungsprozesses?

- (b) (4 Punkte) Erläutern Sie anhand einer geometrischen Darstellung oder der arithmetischen Formeln der Verfahren den Unterschied zwischen Momentum und Nesterov-Momentum.