

מוס' שאלון - 572

14

בגדי 2025

87 מז' מועד

סמסטר 2025 ב

22961 / 4

"ח בתמוז תשפ"ה

שאלון בחינת גמר

22961 - למידה عمוקה

משך בחינה: 3 שעות

בשאלון זה 10 עמודים

מבנה הבחינה:

בחינה 20 שאלות רב ברירה.

יש לענות על כלן.

לכל שאלה יש רק תשובה אחת נכונה.

הקפידו לנצל את הזמן נכון, קר שטספיקו לענות על כל השאלות.

את התשובות עליהם לסמך על גבי גילוי התשובות לשאלות הרוב ברירה

שבגביה מחברת הבחינה.

בסוף הבחינה מצורף ריכוז של נוסחים והגדרות לשימושם.

חומר עזר:

כל חומר עזר אסור לשימוש.

בהצלחה !!!

**איןכם חייבים
להחזיר את השאלון לאוניברסיטה הפתוחה**

שאלה 1

כיצד נקרא השלב בלמידה שבו האלגוריתם עובר פעם אחת על כל הדוגמאות בקבוצת האימון?

- א. Epoch
- ב. Minibatch
- ג. Iteration
- ד. Decay

שאלה 2

בහינתן רשת נוירוניים عمוקה לסיווגRB-מחלקי לעשר קטגוריות (0-9), עם שכבת פלט מסווג softmax, כיצד נכון לעריך את דיוק הסיווג (Accuracy) של המודל?

- א. לחשב את ממוצע הערכים בשכבה ה-softmax.
- ב. לבדוק האם הערך המקסימלי של ה-softmax גבוהה ממספר מסוים (למשל 0.5).
- ג. להשוות את האינדקס של הערך הגבוה ביותר ב-softmax למחלקה האמיתית, ולהחשב את שיעור ההתאמות מכלל הדוגמאות.
- ד. להשוות את האינדקס של הערך הנמוך ביותר ב-softmax למחלקה האמיתית, ולהחשב את שיעור ההתאמות מכלל הדוגמאות.

שאלה 3

במהלך אימון רשת נוירוניים באמצעות SGD (Stochastic Gradient Descent) לעיתים נעשה שימוש בטכניקת תזמון קצב הלמידה (learning rate scheduling). מהי הסיבה המרכזית לשימוש בתזמון כזה?

- א. כדי לשפר את מהירות ההתקנסות על ידי שמירה על קצב למידה גבוה לאורך כל תהליך האימון.
- ב. כדי לאפשר קפיצות קטנות יותר במרקח הפרמטרים בשלבים מתקדמים של האימון, ולהתכנס לפחות מינימום טוב יותר.
- ג. כדי להאיץ את האימון בכך שקצב הלמידה גדול עם כל epoch.
- ד. כדי למנוע מה-gradients להפוך לקטנים מדי לאורך זמן ולשמור קצב שינוי קבוע.
- ה. תשובה א', ב', ג' ו-ד' אינן נכונות.

שאלה 4

כיצד משפייע מומנטום על עדכון הפרמטרים ברשת ניירונים?

- א. הוא מאייך את שינוי הפרמטרים כאשר הגרדיאנטים הולכים וקטנים, כדי להימנע ממצבים בהם הלמידה מאטה באזוריים שטוחים של פונקציית המחיר.
- ב. הוא מחזק תנעوت עקבית בכיוון מסוים ומחליש תנודות בכיוונים אחרים.
- ג. הוא מפסיק את הגרדיאנט כאשר כיוון הירידה משתנה בין אפוקים (epochs) עוקבים.
- ד. הוא מגביר את קצב הלמידה ככל שייתר אפוקים (epochs) נצברים.
- ה. תשובה א', ב', ג' ו-ד' נכונות.
- ו. תשובה א', ב', ג' ו-ד' אינם נכונות.

שאלה 5

מהו היתרונות המרכזי של שימוש באופטימייזר עם קצב למידה אדפטיבי, כמו Adam או RMSprop?

- א. הוא מבטיח התכנסות מהירה יותר מכל אופטימייזר אחר, ללא תלות בנתונים או בארכיטקטורת הרשת.
- ב. הוא מתאים את קצב הלמידה באופן דינמי לכל פרמטר, בהתאם להיסטוריה הגרדיאנטים שלו.
- ג. הוא שומר על קצב למידה נמוך בתחילת האימון, ומעלה אותו באופן הדרגי לאורך האפוקים (epochs).
- ד. הוא מבצע נרמול של כל גרדיאנט לפרמטר כך שלכלם תהיה השפעה שווה על העדכון.
- ה. תשובה א', ב', ג' ו-ד' נכונות.
- ו. תשובה א', ב', ג' ו-ד' אינם נכונות.

שאלה 6

במהלך עיבוד תמונה באמצעות רשת ניירונים קוונבולוציונית (CNN) נעשה שימוש בגרעין קוונבולוציה בגודל 3×3 כדי לזהות קצוטות בתמונה. איזה מהגרעינים הבאים צפוי לזהות **קצוטות אופקיים** (horizontal) בזרה היילה ביותר?

$$\begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix} \text{ א.}$$

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & -1 & 0 \end{bmatrix} \text{ ב.}$$

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix} \text{ ג.}$$

$$\begin{bmatrix} -1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & -1 \end{bmatrix} \text{ ד.}$$

שאלה 7

בailo תנאים שימוש ב - Transfer Learning בלמידה عمוקה נחassoc לאפקטיבי?
בחrho את האפשרות המדויקת ביותר :

- א. Transfer Learning אפקטיבי כאשר שכבות סופיות של רשות מאומנת מראש נלמדות מחדש (fine-tuning) כיון שהן מייצגות תכונות גנריות שמתאימות לרוב המשימות.
- ב. Transfer Learning מתאים בעיקר למשימות שבן יש נתוני אימון בכמות גדולה, כיון שהוא מחזק את הלמידה של המודל המקורי מבליל לשנות את המבנה שלו.
- ג. Transfer Learning נשען על ההנחה שכבות מוקדמות לומדות ייצוגים כלליים הניתנים למיחזור, בעוד שכבות מאוחרות מותאמות למשימה הספציפית ולכך מחליפים או מעודכנים אותן.
- ד. Transfer Learning יכול רק כאשר הארכיטקטורה של המודל המקורי והמודל החדש זהה לחלוטין, כולל מספר שכבות ופרמטרים.

שאלה 8

איזה מה הבאים אינו נחassoc ליתרונו של גישת Transfer Learning ?

- א. מאפשרת לאמן מודלים מדויקים גם כאשר יש כמות קטנה של דוגמאות מותигות.
- ב. חוסכת בזמן חישוב על ידי שימוש במודל קיים שאומן מראש.
- ג. משפרת את ביצועי המודל כאשר תחום המקור שונה מתחום היעד.
- ד. מאפשרת לבצע fine-tuning של שכבות ספציפיות במקום לאמן את כל המודל מחדש.

שאלה 9

נתונות הטענות הבאות בנוגע לרשומות عمוקות לעומת רשומות שטוחות :

- 1. רשומות عمוקות מסוגלות ללמידה ייצוגים היררכיים, כך שכבות מוקדמות לצדדי תבניות פשוטות והעמוקות למדות מבנים מורכבים יותר.
- 2. ככל שמספר השכבות גדול, כך עלות גם הדרישות לזכור ולחישוב, ולעתים נדרש אופטימיזציה מיוחדת כדי להציג את האימון.
- 3. ככל שהרשת عمוקה יותר, כך מוגבר שיפור ביצועי המודל, בזכות יכולת ההבעה הגבוהה.
- 4. בעיות כמו גרדיאנט מתאפס או מתפוץ נפוצות יותר ברשומות عمוקות ודורשות לעיתים שימוש בשיטות כמו *Batch Normalization* או *Residual Connections*.

בחrho בתשובה הנכונה ביותר :

- א. אף טענה אינה נכונה.
- ב. רק טענה אחת נכונה.
- ג. רק שתי טענות נכונות.
- ד. רק שלוש טענות נכונות.
- ה. כל ארבע הטענות נכונות.

שאלה 10

אם רשות נוירונים בעלת שכבה חבויה אחת יכולה לקרב כל פונקציה רציפה בדיקן רצוננו (ע"פ משפט הקירוב האוניברסלי), מדוע בכלל זאת משתמשים ברשותות עמוקות?

בחרו בתשובה הנכונה ביותר:

- א. למרות קיומה של רשות כזו תיאורטית, אלגוריתמים של למידה לא תמיד מצליחים למצוא את הפרמטרים הנכונים באימון בפועל.
- ב. רשותות שטוחות דורשות לרוב מספר עצום של נוירונים כדי לקרב פונקציות מורכבות.
- ג. עומק הרשות מאפשר לה לבנות ייצוגים היררכיים של תכונות, מה שחשוב במיוחד בתחוםים כמו ראייה ממוחשבת או עיבוד שפה.
- ד. תשובה א', ב', ו- ג' נכונות.
- ה. תשובה א', ב', ו- ג' אינן נכונות.

שאלה 11

נתונה רשות הכוללת שתי שכבות: שכבה חבויה אחת עם 10 נוירונים ושכבת פלט אחת לביצוע סיווג בינארי. הפלט לרשות הוא תמונה בגודל 16×16 (כלומר 256 פיקסלים). כמה פרמטרים יש לרשות זו בסך הכל?

- א. 2,561
- ב. 2,570
- ג. 2,581
- ד. 2,691
- ה. 2,571
- ו. תשובה א', ב', ג' ד' ו-ה אינן נכונות.

שאלה 12

רשותFully Connected Feedforward עם שכבה אחת בלבד (שכבת הפלט) עשויה שימוש בפונקציית אקטיבציה מסוג sigmoid לצורך סיווג ביןארי לשתי קטגוריות: "0" או "1". הפלט של הרשות מפורש כיחסות שחקלט שייך למחלקה החיוונית ("1"). המודל מאומן באמצעות עקרון הסבירות המירבית (Maximum Likelihood). איזו מהטענות הבאות מתארת באופן המדויק ביותר את השפעת תופעת הרויה (saturation) של פונקציית ה- sigmoid על הגרדיאנט ועל תהליכי הלמידה?

- א. בעיית הרויה קיימת רק כאשר המודל חוזה נכון, ולכן אינה פוגעת בתהליכי הלמידה.
- ב. הגרדיאנט עלול להיות קטן מדי ולחסום את הלמידה בכל פעם שמרתחשת רויה, בין אם המודל חוזה נכון ובין אם לא.
- ג. Cross Entropy מבטלת את השפעת רויה הסיגמוואיד על הגרדיאנט של פונקציית העלות, כאשר המודל חוזה נכון וגם במקרה שלא.
- ד. תשובה א', ב', ו-ג' אינן נכונות.

שאלה 13

נתונה רשת עצבית מסווג Fully Connected Feedforward הכוללת שכבה חבויה אחת, כאשר כל הנוירונים בשכבה זו משתמשים בפונקציית האקטיבציה **sigmoid**:

$$\sigma(x) = \frac{1}{1+e^{-x}}$$

כעת מחליפים את פונקציית האקטיבציה בשכבה החבויה לפונקציות **הטנגנס ההיפרבולי**:

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

האם ניתן להתאים את פרמטרי הרשות (משקלים והטויות) כך שהרשota החדש שבח פונקציות האקטיבציה בשכבה החבויה הן \tanh תהייה **סקולה**, כלומר, תניב את אותה פונקציית מיופי, לרשות המקורית שבה נעשה שימוש ב- σ ? הניחו כי בשתי הרשותות, פונקציית האקטיבציה ביחידת הפלט היא פונקציית זהות (Identity).

- א. ניתן להתאים - לא צפוי שינוי מהותי בפרמטרי הרשות, משום שלשתי הפונקציות בתחום ערכים דומים.
- ב. ניתן להתאים - נקבע פיא שניים את המשקלים וההטיה בכניסה לכל \tanh , נקבע פיא שניים את המשקל בモוצא כל \tanh ונתאים את ערך ההטיה בכניסה לשכבות הפלט.
- ג. ניתן להתאים - נגדיל פיא שניים את המשקלים וההטיה בכניסה לכל \tanh , נגדיל פיא שניים את המשקל ב모וצא כל \tanh ונתאים את ערך ההטיה בכניסה לשכבות הפלט.
- ד. ניתן להתאים - יידרשו משקלים שליליים בלבד, כי \tanh מחזירה גם ערכים שליליים, בנגדוד ל-**sigmoid**.
- ה. לא ניתן להתאים את פרמטרי הרשות כדי לקבל שקיולות.
- ו. תשובה א', ב', ג', ד ו-ה אין נכונות.

שאלות 14,15:

נתונה פונקציית מחיר (cost function) ריבועית מהצורה:

$$f(\mathbf{w}) = \frac{1}{2} \mathbf{w}^T Q \mathbf{w} + \mathbf{b}^T \mathbf{w}$$

כאשר $Q \in \mathbb{R}^{2 \times 2}$ היא מטריצה סימטרית עם ערכים עצמיים 5 ווקטורים עצמיים מתאימים $\mathbf{w}^* = [1, -1]^T$. הנקודה $\mathbf{v}_1 = [1, 2]/\sqrt{5}$, $\mathbf{v}_2 = [-2, 1]/\sqrt{5}$ מביאה למינימום את פונקציית המחיר נתונה.

שאלה 14

הוקטור \mathbf{b} נתון ע"י

- . $\mathbf{b} = [-1, 5.5]^T$.א.
- . $\mathbf{b} = [1, -5.5]^T$.ב.
- . $\mathbf{b} = [-2, 11]^T$.ג.
- . $\mathbf{b} = [2, -11]^T$.ד.
- . $\mathbf{b} = [-0.5, 2.75]^T$.ה.
- .תשובות א', ב', ג', ד' זה אין נכון.

שאלה 15

מוסיפים לפונקציית המחיר איבר רגולרייזציה מסוג Ridge :

$$f(\mathbf{w}) = \frac{1}{2} \mathbf{w}^T Q \mathbf{w} + \mathbf{b}^T \mathbf{w} + \frac{\lambda}{2} \|\mathbf{w}\|_2^2$$

עם פרמטר $\lambda = 10$. נקודת המינימום \mathbf{w}_{reg} של פונקציית המחיר החדשה עם איבר הרגולרייזציה נתונה ע"י:

- . $\mathbf{w}_{reg} \approx [0.27, -0.47]^T$.א.
- . $\mathbf{w}_{reg} \approx [-0.27, 0.47]^T$.ב.
- . $\mathbf{w}_{reg} \approx [0.47, -0.27]^T$.ג.
- . $\mathbf{w}_{reg} \approx [-0.47, 0.27]^T$.ד.
- . $\mathbf{w}_{reg} \approx [-0.27, 0.27]^T$.ה.
- . $\mathbf{w}_{reg} \approx [0.27, -0.27]^T$.ו.

שאלה 16

בעת שימוש בשיטת Gradient Descent לאופטימיזציה של פונקציית עלות קמורה וחלקה, נמצא כי מטריצת ההסיאן (hessian) של הפונקציה בעלת condition number גבוהה. איזו מהפעולות הבאות אינה כפופה לשפר את קצב התכנסות?

- ביצוע נרמול (normalization) או סטנדרטיזציה של משתני הקלט, כך שכל משתנה יהיה בטווח דומה או בעל סטיית תקן אחידה.
- שימוש באופטימיזר כמו **Adam** שמתאים את גודל העדכון בכל כיוון על פי הגרדיינטים ההיסטוריים של הפונקציה.
- שינוי קווארדינטות שמאזן את קנה המידה של הциונים השונים.
- הוספת Dropout בשכבות החבויות כדי להקטין תלות בין יחידות ולשפר הכללה.

שאלה 17

בשני מודלים של רשתות נוירונים عمוקות נוסף איבר רגולרייזציה שונה לפונקציית העלות:

- במודל א' נעשה שימוש ברגולרייזציה מסוג L^2 .
- במודל ב' נעשה שימוש ברגולרייזציה מסוג L^1 .

איזו מהטענות הבאות אינה נכונה בנוגע להבדלים בין שני סוגי הרגולרייזציה ולהשפעתם על תהליכי האופטימיזציה ברשתות عمוקות?

- רגולרייזציה L^2 נוטה להקטין את ערכי המשקלים אך לרוב תשאיר אותם שונים מאפס.
- רגולרייזציה L^1 מעודדת דילולות (sparsity), כלומר הרבה משקלים יהיו בדיק אפס.
- שימוש ברגולרייזציה L^1 ברשתות عمוקות מצריך אופטימיזציה איטרטיבית, כיון שאין פתרון אנלטי סגור (closed-form solution).
- שימוש ברגולרייזציה L^2 ברשתות عمוקות אינו מצריך אופטימיזציה איטרטיבית, כיון שיש פתרון אנלטי סגור (closed-form solution).

שאלה 18

איזו מהthesesות הבאות אינה נכונה לגבי השימוש ב- Dropout במהלך אימון רשתות נוירונים عمוקות?

- Dropout מונע התחשבות יתר על ידי השמטת אקראית של נוירונים בשכבות במהלך האימון.
- Dropout משפר את יכולת ההכללה של הרשת על ידי ייצור "תת-רשתות" שונות בכל איטרציה.
- Dropout מתאים רק בשכבות החבויות ולא ניתן להפעיל אותו בשכבות הקלט.
- בשלב הבדיקה (inference) לא מבצעים Dropout אלא משתמשים בכל הנוירונים עם משקלים מותאמים.

שאלה 19

מהם ה יתרונות של שימוש בשכבות קונבולוציה (Convolutional Layers) לעומת שכבות צפופות (Fully Connected) במשימות חזותיות?

בחרו איזו מההתשובות הבאות אינה נכונה:

- א. שכבות קונבולוציה עושות שימוש בהקשר מרחבי.
- ב. שכבות קונבולוציה מאפשרות אינואריאנטיות להזזה.
- ג. לשכבות קונבולוציה יש הרבה פחות פרמטרים.
- ד. שכבות קונבולוציה דורשות יותר פרמטרים מאשר שכבות צפופות כדי לתפוס את המורכבות של התמונה.

שאלה 20

רשת קונבולוציה מקבלת קלט טנזור קלט ממימד 128×128 העובר דרך שכבת קונבולוציה בעלת 16 פילטרים עם גודל גרעין 3×3 , stride 1×3 , ולא ריפוד באפסים. מה יהיה גודל טנзор הפלט?

בחרו את התשובה הנכונה:

- א. $16 \times 126 \times 124$.
- ב. $16 \times 126 \times 42$.
- ג. $16 \times 42 \times 126$.
- ד. $16 \times 125 \times 41$.
- ה. תשובה א', ב', ג' ו-ד' אינן נכונות.

ב ה צ ל ח ח !

רכיב הגדרות ונוסחאות:

$$\text{sigmoid}(x) = \frac{1}{1 + e^{-x}}$$

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

Derivatives:

$$\frac{\partial}{\partial \mathbf{w}}(\mathbf{w}^T \mathbf{x}) = \mathbf{x}$$

$$\frac{\partial}{\partial \mathbf{w}}(\mathbf{w}^T A \mathbf{w}) = (A + A^T)\mathbf{w}$$

$$\frac{\partial^2}{\partial \mathbf{w} \partial \mathbf{w}^T}(\mathbf{w}^T A \mathbf{w}) = A + A^T$$

The **directional derivative** in the direction of \mathbf{u} where \mathbf{u} is a unit vector:

$$\frac{\partial}{\partial \alpha} f(\mathbf{x} + \alpha \mathbf{u}) \Big|_{\alpha=0} = \mathbf{u}^T \nabla_{\mathbf{x}} f(\mathbf{x})$$

The **second directional derivative** in the direction of \mathbf{u} where \mathbf{u} is a unit vector:

$$\frac{\partial^2}{\partial \alpha^2} f(\mathbf{x} + \alpha \mathbf{u}) \Big|_{\alpha=0} = \mathbf{u}^T \mathbf{H}(\mathbf{x}) \mathbf{u}$$

where $\mathbf{H}(\mathbf{x})$ is the Hessian matrix: $H(\mathbf{x})_{i,j} = \frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{x})$.

A second-order Taylor approximation around the point $\mathbf{x}^{(0)}$:

$$f(\mathbf{x}) \approx f(\mathbf{x}^{(0)}) + (\mathbf{x} - \mathbf{x}^{(0)})^T \nabla_{\mathbf{x}} f(\mathbf{x})|_{\mathbf{x}=\mathbf{x}^{(0)}} + \frac{1}{2} (\mathbf{x} - \mathbf{x}^{(0)})^T \mathbf{H}(\mathbf{x}^{(0)}) (\mathbf{x} - \mathbf{x}^{(0)})$$