



מספר תעודת זהות
יש לרשום את כל תשע הספרות

האוניברסיטה
הפתוחה



יש להדביק כאן את
מדבקת הנבחן

י"ח בתמוז תשפ"ה

מס' שאלון - 572
14 ביולי 2025

מס' מועד 87

סמסטר 2025ב

22961 / 4

שאלון בחינת גמר
22961 - למידה עמוקה

משך בחינה: 3 שעות

בשאלון זה 10 עמודים

מבנה הבחינה:

בבחינה 20 שאלות רב ברירה.

יש לענות על כולן.

לכל שאלה יש רק תשובה אחת נכונה.

הקפידו לנהל את הזמן נכון, כך שתספיקו לענות על כל השאלות.

את התשובות עליכם לסמן על גבי גיליון התשובות לשאלות הרב ברירה

שבגב מחברת הבחינה.

בסוף הבחינה מצורף ריכוז של נוסחאות והגדרות לשימושכם.

חומר עזר:

כל חומר עזר אסור לשימוש.

בהצלחה !!!

אינכם 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

בהינתן רשת נוירונים עמוקה לסיווג רב-מחלקתי לעשר קטגוריות (0-9), עם שכבת פלט מסוג softmax, כיצד נכון להעריך את דיוק הסיווג (Accuracy) של המודל?

- א. לחשב את ממוצע הערכים בשכבת ה-softmax.
- ב. לבדוק האם הערך המקסימלי של ה-softmax גבוה מסף מסוים (למשל 0.5).
- ג. להשוות את האינדקס של הערך הגבוה ביותר ב-softmax למחלקה האמיתית, ולחשב את שיעור ההתאמות מכלל הדגימות.
- ד. להשוות את האינדקס של הערך הנמוך ביותר ב-softmax למחלקה האמיתית, ולחשב את שיעור ההתאמות מכלל הדגימות.

שאלה 3

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

- א. כדי לשפר את מהירות ההתכנסות על ידי שמירה על קצב למידה גבוה לאורך כל תהליך האימון.
- ב. כדי לאפשר קפיצות קטנות יותר במרחב הפרמטרים בשלבים מתקדמים של האימון, ולהתכנס למינימום טוב יותר.
- ג. כדי להאיץ את האימון בכך שקצב הלמידה גדל עם כל epoch.
- ד. כדי למנוע מה- gradients להפוך לקטנים מדי לאורך זמן ולשמר קצב שינוי קבוע.
- ה. תשובות א', ב', ג' ו-ד' אינן נכונות.

שאלה 4

כיצד משפיע מומנטום על עדכון הפרמטרים ברשת נוירונים?

- א. הוא מאיץ את שינוי הפרמטרים כאשר הגרדיאנטים הולכים וקטנים, כדי להימנע ממצבים בהם הלמידה מאטה באזורים שטוחים של פונקציית המחר.
- ב. הוא מחזק תנועות עקביות בכיוון מסוים ומחליש תנודות חדות בכיוונים משתנים.
- ג. הוא מאפס את הגרדיאנט כאשר כיוון הירידה משתנה בין אפוקים (epochs) עוקבים.
- ד. הוא מגביר את קצב הלמידה ככל שיותר אפוקים (epochs) נצברים.
- ה. תשובות א', ב', ג' ו-ד' נכונות.
- ו. תשובות א', ב', ג' ו-ד' אינן נכונות.

שאלה 5

מהו היתרון המרכזי של שימוש באופטימיזר עם קצב למידה אדפטיבי, כמו Adam או RMSprop?

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

שאלה 6

במהלך עיבוד תמונה באמצעות רשת נוירונים קונבולוציונית (CNN) נעשה שימוש בגרעין קונבולוציה בגודל 3×3 כדי לזהות קצוות בתמונה. איזה מהגרעינים הבאים צפוי לזהות קצוות אופקיים (horizontal) בצורה היעילה ביותר?

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

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

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

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

שאלה 7

באילו תנאים שימוש ב - Transfer Learning בלמידה עמוקה נחשב לאפקטיבי?
בחרו את האפשרות המדויקת ביותר :

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

שאלה 8

איזה מהבאים **אינו** נחשב ליתרון של גישת Transfer Learning?

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

שאלה 9

נתונות הטענות הבאות בנוגע לרשתות עמוקות לעומת רשתות שטוחות :

1. רשתות עמוקות מסוגלות ללמוד ייצוגים היררכיים, כך ששכבות מוקדמות לוכדות תבניות פשוטות והעמוקות לומדות מבנים מורכבים יותר.
2. ככל שמספר השכבות גדל, כך עולות גם הדרישות לזיכרון ולחישוב, ולעיתים נדרשת אופטימיזציה מיוחדת כדי להאיץ את האימון.
3. ככל שהרשת עמוקה יותר, כך מובטח שיפור בביצועי המודל, בזכות יכולת ההבעה הגבוהה.
4. בעיות כמו גרדיאנט מתאפס או מתפוצץ נפוצות יותר ברשתות עמוקות ודורשות לעיתים שימוש בשיטות כמו Batch Normalization או Residual Connections.

בחרו 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 רשת נוירונים בעלת שכבה חבויה אחת יכולה לקרב כל פונקציה רציפה בדיוק כרצוננו (ע"פ משפט הקירוב האוניברסלי), מדוע בכל זאת משתמשים ברשתות עמוקות?

בחרו בתשובה הנכונה ביותר :

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

שאלה 11

נתונה רשת Fully Connected Feedforward הכוללת שתי שכבות: שכבה חבויה אחת עם 10 נוירונים ושכבת פלט אחת לביצוע סיווג בינארי. הקלט לרשת הוא תמונה בגודל 16×16 (כלומר 256 פיקסלים). כמה פרמטרים יש ברשת זו בסך הכל?

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

שאלה 12

רשת Fully Connected Feedforward עם שכבה אחת בלבד (שכבת הפלט) עושה שימוש בפונקציית אקטיבציה מסוג sigmoid לצורך סיווג בינארי לשתי קטגוריות: "0" או "1". הפלט של הרשת מפורש כהסתברות שהקלט שייך למחלקה החיובית ("1"). המודל מאומן באמצעות עקרון הסבירות המירבית (Maximum Likelihood). איזו מהטענות הבאות מתארת באופן המדויק ביותר את השפעת תופעת הרוויה (saturation) של פונקציית ה-sigmoid על הגרדיאנט ועל תהליך הלמידה?

- א. בעיית הרוויה קיימת רק כאשר המודל חוזה נכון, ולכן אינה פוגעת בתהליך הלמידה.
- ב. הגרדיאנט עלול להיות קטן מדי ולחסום את הלמידה בכל פעם שמתרחשת רוויה, בין אם המודל חוזה נכון ובין אם לא.
- ג. Cross Entropy מבטלת את השפעת רוויית הסיגמואיד על הגרדיאנט של פונקציית העלות, כאשר המודל חוזה נכון וגם כשלא.
- ד. תשובות א', ב', ו-ג' אינן נכונות.

שאלה 13

נתונה רשת עצבית מסוג 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 פונקציות האקטיבציה בשכבה החבויה לפונקציות **הטנגנס ההיפרבולי**:

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

האם ניתן להתאים את פרמטרי הרשת (משקלים והטיות) כך שהרשת החדשה שבה פונקציות האקטיבציה בשכבה החבויה הן \tanh תהיה **שקולה**, כלומר, תניב את אותה פונקציית מיפוי, לרשת המקורית שבה נעשה שימוש ב- σ ? הניחו כי בשתי הרשתות, פונקציית האקטיבציה ביחידת הפלט היא פונקציית הזהות (Identity).

- א. ניתן להתאים - לא צפוי שינוי מהותי בפרמטרי הרשת, משום שלשתי הפונקציות תחום ערכים דומה.
- ב. ניתן להתאים - נקטין פי שניים את המשקלים וההטיות בכניסה לכל \tanh , נקטין פי שניים את המשקל במוצא כל \tanh ונתאים את ערך ההטיות בכניסה לשכבת הפלט.
- ג. ניתן להתאים - נגדיל פי שניים את המשקלים וההטיות בכניסה לכל \tanh , נגדיל פי שניים את המשקל במוצא כל \tanh ונתאים את ערך ההטיות בכניסה לשכבת הפלט.
- ד. ניתן להתאים - יידרשו משקלים שליליים בלבד, כי \tanh מחזירה גם ערכים שליליים, בניגוד ל-sigmoid.
- ה. לא ניתן להתאים את פרמטרי הרשת כדי לקבל שקילות.
- ו. תשובות א', ב', ג', ד ו-ה אינן 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 $\lambda_1 = 20, \lambda_2 = 5$ ווקטורים עצמיים מתאימים $\mathbf{v}_1^T = [1, 2]/\sqrt{5}, \mathbf{v}_2^T = [-2, 1]/\sqrt{5}$. הנקודה $\mathbf{w}^{*T} = [1, -1]$ מביאה למינימום את פונקציית המחיר נתונה.

שאלה 14

הוקטור \mathbf{b} נתון ע"י

א. $\mathbf{b} = [-1, 5.5]^T$

ב. $\mathbf{b} = [1, -5.5]^T$

ג. $\mathbf{b} = [-2, 11]^T$

ד. $\mathbf{b} = [2, -11]^T$

ה. $\mathbf{b} = [-0.5, 2.75]^T$

ו. תשובות א', ב', ג', ד' ו-ה אינן נכונות.

שאלה 15

מוסיפים לפונקציית המחיר איבר רגולריזציה 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 של פונקציית המחיר החדשה עם איבר הרגולריזציה נתונה ע"י:

א. $\mathbf{w}_{reg} \approx [0.27, -0.47]^T$

ב. $\mathbf{w}_{reg} \approx [-0.27, 0.47]^T$

ג. $\mathbf{w}_{reg} \approx [0.47, -0.27]^T$

ד. $\mathbf{w}_{reg} \approx [-0.47, 0.27]^T$

ה. $\mathbf{w}_{reg} \approx [-0.27, 0.27]^T$

ו. $\mathbf{w}_{reg} \approx [0.27, -0.27]^T$

שאלה 16

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

- 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 הגרדיאנטים ההיסטוריים של הפונקציה.
- שינוי קואורדינטות שמאזן את קנה המידה של הכיוונים השונים.
- הוספת Dropout בשכבות החביות כדי להקטין תלות בין יחידות ולשפר הכללה.

שאלה 17

בשני מודלים של רשתות נוירונים עמוקות נוסף איבר רגולריזציה שונה לפונקציית העלות:

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

איזו מהטענות הבאות **אינה נכונה** בנוגע להבדלים בין שני סוגי הרגולריזציה ולהשפעתם על תהליך האופטימיזציה ברשתות עמוקות?

- רגולריזציה L^2 נוטה להקטין את ערכי המשקלים אך לרוב תשאיר אותם שונים מאפס.
- רגולריזציה L^1 מעודדת דלילות (sparsity), כלומר הרבה משקלים יהיו בדיוק אפס.
- שימוש ברגולריזציה L^1 ברשתות עמוקות מצריך אופטימיזציה איטרטיבית, כיוון שאין פתרון אנליטי סגור (closed-form solution).
- שימוש ברגולריזציה L^2 ברשתות עמוקות אינו מצריך אופטימיזציה איטרטיבית, כיוון שיש פתרון אנליטי סגור (closed-form solution).

שאלה 18

איזו מהתשובות הבאות **אינה נכונה** לגבי השימוש ב-Dropout במהלך אימון רשתות נוירונים עמוקות?

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

שאלה 19

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

בחרו איזו מהתשובות הבאות **אינה נכונה** :

- א. שכבות קונבולוציה עושות שימוש בהקשר מרחבי.
- ב. שכבות קונבולוציה מאפשרות אינוואריאנטיות 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 כקלט טנזור קלט ממימד 128×128 העובר דרך שכבת קונבולוציה בעלת 16 פילטרים עם גודל גרעין 3×5 , 1×3 stride, וללא ריפוד באפסים. מה יהיה גודל טנזור 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. תשובות א', ב', ג' ו-ד' אינן נכונות.

ב ה צ ל ח ה !

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

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

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:

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

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

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

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 $\mathbf{x}^{(0)}$:

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