

כ"ז באב תשפ"ה

סמסטר 2025 ב

22961 / 4

572 מ"ש שאלון - 21  
באוגוסט 2025

93 מ"ש מועד

שאלון בחינת גמר

22961 - למידה عمוקה

משך בחינה: 3 שעות

בשאלון זה 11 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 عليיכם לסמן על גבי כרטיס המחשב שמצוף לשאלון.  
בסוף הבחינה מצורף ריכוז של נוסחאות והגדרות לשימושכם.

חומר עזר:

כל חומר עזר אסור לשימוש.

בהצלחה !!!

החזירו  
למשגיח את השאלון  
וכל עזר אחר שקיבלתם בתוך מחברת התשובות

### שאלה 1

איזה מהשיטות הבאות נחשבת לשיטת רגולרייזציה (regularization)?

- א. Dropout
- ב.  $L^2$  penalty
- ג. Early stopping
- ד. Data Augmentation
- ה. תשובות א', ב', ג' ו-ד' נכונות.
- ו. תשובות א', ב', ג' ו-ד' אינם נכונות.

### שאלה 2

באיזה מצב השתמש בפונקציית אקטיבציה מסוג sigmoid? בחרו בתשובה הנכונה ביותר.

- א. כאשר אנו מבצעים סיווג רב-מחלكتי.
- ב. כאשר אנו מבצעים רגרסיה לינארית.
- ג. כאשר יש לנו משימה של סיווג ביןארי.
- ד. כאשר נרצה להחליף את softmax.
- ה. תשובות א', ב', ג' ו-ד' נכונות.
- ו. תשובות א', ב', ג' ו-ד' אינם נכונות.

### שאלה 3

בנהנעה שיש ברשותכם רשות מאומנת מראש על מיליון תמונות של אובייקטים כלליים (כגון ImageNet), ואתם מעוניינים לאמן מודל חדש לזיהוי פטולוגיות רפואיות בצילומי רנטגן, אך במשימה החדשיה יש מספר מצומצם של דוגמאות אימון, איך אסטרטגיית Transfer Learning תיחסב לאפקטיבית ביותר? בחרו את האפשרות המדויקת ביותר.

- א. לאמן את כל שכבות הרשות מחדש (training from scratch) כיון שהמשימה שונה באופן מהותי מהמשימה המקורי ודורשת ייצוגים מותאמים.
- ב. לקבע (freeze) את השכבות הראשונות ולאמן מחדש את השכבות העליונות (fine-tuning) כיון שכבות מוקדמות לומדות ייצוגים כלליים שניתנים להעברה גם למשימות שונות, וההתאמאה נעשית בשכבות הגבוהות.
- ג. לקבע (freeze) את כל שכבות הרשות המקורי ולהוסיף שכבה סופית חדשה שתאומן מלהתחילה.
- ד. להחליף את כל השכבות חוץ מהראשונה, כדי לשמור רק את השלב הראשוני של עיבוד התמונה, המשותף לרוב התחומים.

#### שאלה 4

ברשותות קונבולוציה (CNN), געשה שימוש בשכבות Pooling כחלק מהארQUITטורה של הרשת. מהו התפקיד המרכזי של שכבות אלו בתהליך הלמידה?

##### בחרו את האפשרות המדויקת ביותר.

- א. מנרמלות את ערכי הקלט כדי לשפר את יציבות האימון.
- ב. מוסיפות עומק לרשת ובכך מאפשרות למודל ללמידה ייצוגים מורכבים יותר.
- ג. מאייצות את תהליך האופטימיזציה באמצעות שינוי הגרדייאנטים בשכבות הקונבולוציה.
- ד. מצמצמות את המימדים המרחביים של הנתונים, מפחיתות את מספר הפרמטרים, ומאפשרות למודל לזהות תכונות באופן חסין לשינויים במקומות (translation invariance).
- ה. משפרות את דיקן המודול על ידי הקטנת שגיאות הניבוי באופן ישיר.

#### שאלה 5

במהלך יישום של אלגוריתם מורד הגרדייאנט (Gradient Descent) לאופטימיזציה של פונקציית עלות שאינה קמורה, איזו מהטענות הבאות מדויקת ביותר?

- א. האלגוריתם מובטח להגיע למינימום הגלובלי, ללא תלות במבנה הפונקציה.
- ב. האלגוריתם עלול להתכנס למינימום מקומי או לנקודת אוכף (saddle point), והתוצאה העשויה להיות תלולה בנקודת ההתחלתה.
- ג. קצב הלמידה (learning rate) משפייע רק על מהירות ההתכנסות, אך אינו משפייע על איכות הפתרון הסופי.
- ד. כל עוד מתבצעים עדכונים לפי הגרדייאנט, מובטחת ההתכנסות לפתרון יציב.
- ה. אלגוריתם מורד הגרדייאנט מתאים רק לפונקציות קמורות. עבור פונקציות לא קמורות אין דרך להשתמש בו.
- ו. בין התשובות א', ב', ג', ד' ו-ה ישנה יותר מתשובה אחת נכונה.

## שאלה 6

נתונה רשת נוירונים מסווג Fully Connected Feedforward, הכוללת שכבה קלט, שכבה חבויה אחת, ושכבה פלט אחת עם נוירון אחד לצורך סיווג ביןארי. בכל שכבה, כל נוירון מקבל קלט מכל הנוירונים בשכבה הקודמת וכול גם הטיה (bias). מימד הקלט הוא  $20 \times 20$ . ידוע כי מספר הפרמטרים הכלול (כולל משקלים והטיות) בראשת זו הוא 5,629. מהו מספר הנוירונים בשכבה החבויה?

- א. 12.
- ב. 13.
- ג. 14.
- ד. 15.
- ה. לא ניתן לקבוע מהתנאים.
- ו. תשובה א', ב', ג' ו-ה אינן נכונות.

## שאלה 7

במודלים של למידה מפוקחת (supervised learning), מקובל להשתמש ברגולרייזציה מסווג  $L^1$  או רגולרייזציה מסווג  $L^2$  כדי להפחית סיכון להתקאת יתר (overfitting). מהו ההבדל המרכזי בין שתי שיטות הרגולרייזציה? **בחרו את האפשרות המדויקת ביותר.**

- א. רגולרייזציה מסווג  $L^2$  נוטה לאפס חלק מהמשקלים לחלוין, ולכן יוצרת פתרונות דילילים (sparse).
- ב. רגולרייזציה מסווג  $L^1$  שומרת על כל הפרמטרים קטנים אך לא מאפסת אותם.
- ג. רגולרייזציה מסווג  $L^1$  שומרת את כל הפרמטרים בתוצאה, ולכן אינה מתאימה לצמצום מספר הפיצירים הפעילים במודל.
- ד. רגולרייזציה מסווג  $L^1$  תמיד עדיפה על רגולרייזציה מסווג  $L^2$ , במיוחד ברשתות נוירונים עמוקות.
- ה. תשובה א', ב', ג' ו-ד' נכונות.
- ו. תשובה א', ב', ג' ו-ד' אינן נכונות.

## שאלה 8

מהי המטרה העיקרית של שימוש בטכניקת Early Stopping במהלך אימון רשתות נירונים?  
**בחרו את התשובה הנכונה ביותר.**

- א. להפסיק את האימון כאשר השיפור בביוצוי המודל על קבוצת האימון נעצר או מתחילה להיפגע, כדי לחסוך בזמן חישוב.
- ב. להפסיק את האימון כאשר הביצועים על קבוצת הולידציה מפסיקים להשתפר ומתחלים להידרדר, במטרה למנוע התאמות יתר (overfitting).
- ג. להפסיק את האימון לאחר מספר אפוקים קבוע מראש, ללא קשר לביצועי המודל על קבוצות הנתונים.
- ד. להפסיק את האימון כאשר שגיאת האימון יורדת מתחת למספר מסוים ומתייצבת.
- ה. להפסיק את האימון כשפונקציית loss משתפרת בהתמדה לאורך מספר אפוקים רצופים.
- ו. תשובה א', ב', ג' ו-ד' אינה נכונה.

## שאלה 9

מהו התפקיד המרכזי של פונקציית האקטיבציה (Activation function) ברשת נירונים?

- א. לנרמל או לקודד את נתוני הקלט לפני המעבר לשכבות הolute.
- ב. לאפשר לרשת ללמידה קשרים לא לינאריים בין הקלט לפולט.
- ג. לאפשר למודל להתכנס במהירות במהלך תהליך האימון.
- ד. להפחית את הסיכון להתקנת יתר (overfitting) באמצעות רגולרייזציה פנימית.
- ה. תשובה א', ב', ג' ו-ד' אינה נכונה.

## שאלה 10

במהלך אימון רשתות נוירוניים عمוקות במיוחד (למשל, בעלות כ-100 שכבות) למשימות סיווג ביןארי, נעשה שימוש בפונקציית Sigmoid בשכבה הפלט בשילוב של פונקציות אקטיבציה מסווג  $\text{tanh}$  ו-ReLU בשכבות הפנימיות. לאחר epoch אחד בלבד, מתרברר כי בשכבות מסוימות המשקלים מפסיקים להתקדם, אף שהרשת לא התקנסה. בדיקה מעמיקה מעלה כי הגרדיינטס באוטן שכבות נעלמים כמעט לגמרי כבר בשלבים מוקדמים של האימון, בעוד שפונקציית ההפסד (Loss) נשארת בטוחה סביר.

בהת总算ב בכך, אילו מהשינויים האפשריים הבאים עשויים לסייע במצטום הבעיה?

1. להגדיל את גודל קבוצת האימון.
2. להחליף את פונקציות ReLU בפונקציות Leaky ReLU.
3. להוסיף Batch Normalization לפני כל פונקציית אקטיבציה.
4. להגדיל את קצב הלמידה (learning rate).

**בחרו את האפשרות הנכונה ביותר :**

- א. רק שניים אחד מトוך הארבעה עשויים לסייע במצטום הבעיה.
- ב. רק שניים מトוך הארבעה עשויים לסייע במצטום הבעיה.
- ג. שלושה שינויים מトוך הארבעה עשויים לסייע במצטום הבעיה.
- ד. כל ארבעת השינויים המוצעים עשויים לסייע במצטום הבעיה.
- ה. אף אחד מהשינויים המוצעים אינו עשוי לסייע במצטום הבעיה.

## שאלה 11

Dropout היא טכניקת רגולרייזציה נפוצה באימון רשתות נוירוניים.

**בחרו בתשובה המדויקת ביותר .**

- א. Dropout מוביל לדילולות (sparsity) במשקלים המאומנים של הרשות.
- ב. במהלך שלב הבדיקה (test time), Dropout מיושם עם הסתברות היישארות (keep probability) הפוכה.
- ג. ככל שהסתברות היישארות גבוהה יותר, כך האפקט של הרגולרייזציה מתחזק.
- ד. תשובות א' ו-ב' נכונות.
- ה. תשובות א', ב' ו-ג' נכונות.
- ו. תשובות א', ב', ו-ג' אינן נכונות.

### שאלות 12,13

נתונה פונקציית עלות (cost function) Ridge רגולרייזציה מסוג מהצורה:

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

כאשר  $Q \in \mathbb{R}^{2 \times 2}$  היא מטריצה סימטרית עם ערכים עצמיים  $\lambda_1 = 10, \lambda_2 = 2.5$  ווקטורים עצמיים מתאימים  $\mathbf{v}_1^T = [1, 2]/\sqrt{5}, \mathbf{v}_2^T = [-2, 1]/\sqrt{5}$ .

ידוע כי עבור ערך רגולרייזציה  $\lambda = 10$ , פתרון בעיית המינימום הוא

$$\mathbf{w}_{reg} = [\frac{7}{100}, -\frac{4}{25}]^T$$

### שאלה 12

מהו הפתרון  $\mathbf{w}^*$  המוצע את פונקציית העלות כאשר אין איבר רגולרייזציה (כלומר, כאשר  $\lambda = 0$ )?

- .  $\mathbf{w}^{*T} = [1/2, -1/2]$  א.
- .  $\mathbf{w}^{*T} = [-1/2, 1/2]$  ב.
- .  $\mathbf{w}^{*T} = [1/2, 0]$  ג.
- .  $\mathbf{w}^{*T} = [0, -1/2]$  ד.
- .  $\mathbf{w}^{*T} = [1/2, -1/25]$  ה.
- . תשובות א', ב', ג', ד' זה אין נכונות. ו.

### שאלה 13

הוקטור  $\mathbf{b}$  נתון ע"י

- .  $\mathbf{b} = [-0.5, 2.75]^T$  א.
- .  $\mathbf{b} = [1, -5.5]^T$  ב.
- .  $\mathbf{b} = [-1, 5.5]^T$  ג.
- .  $\mathbf{b} = [0.5, -2.75]^T$  ד.
- .  $\mathbf{b} = [-0.25, 1.375]^T$  ה.
- . תשובות א', ב', ג', ד' זה אין נכונות. ו.

### שאלה 14

רשת כוונבולוציה קיבלת כקלט טנзор ממימד  $256 \times 64$  העובר דרך שכבה כוונבולוציה בעלת 32 פילטרים עם גודל גרעין  $10 \times 6$ , stride  $5 \times 3$ , ולא ריפוד באפסים. מה יהיה גודל טנзор הפלט?

**בחרו את התשובה הנכונה:**

- א.  $32 \times 50 \times 20$ .
- ב.  $32 \times 49 \times 20$ .
- ג.  $32 \times 50 \times 19$ .
- ד.  $32 \times 49 \times 19$ .
- ה.  $32 \times 19 \times 49$ .
- ו. תשובות א', ב', ג', ד' ו-ה אינן נכונות.

### שאלה 15

בעת אימון רשת נוירונים באמצעות אלגוריתם SGD, נדרש לבחור גודל מתאים למיני-באץ' (mini-batch). מהם ה יתרונות והחסרונות של שימוש במיני-באץ' קטנים לעומת מיני-באץ' גדולים?

**בחרו את התשובה הנכונה ביותר.**

- א. מיני-באץ' קטנים מוביילים לעדכוני משקל תכופים יותר, עם גרדיאנטים רועשים יותר, מה שיכול לסייע ביציאה ממינימום מקומי אך עשוי להאט את ההתקנסות ולדרוש קצב למידה קטן יותר.
- ב. מיני-באץ' גדולים מפחיתים את רעש הגראדיאנט, מה שתורם להתקנסות יציבה יותר במספר צעדים קטן יותר, אך הם דורשים יותר זיכרון חישובי ועלולים להיתקע במינימום מקומי.
- ג. מיני-באץ' גדולים צורכים פחות משאבי זיכרון וכן הם תמיד עדיפים חישובית.
- ד. בפועל, גודל המיני-באץ' נבחר לעיתים קרובות על פי מגבלות זכרו - GPU ולא רק משיקולי אופטימיזציה תיאורטיים.
- ה. שימוש במיני-באץ' קטנים מוגביר את הסיכון להסתמת יתר (overfitting) וכן יש להימנע מהם.
- ו. בין התשובות א', ב', ג', ד' ו-ה יש יותר מתשובה אחת נכונה.

### שאלה 16

- מהו העיקרי המרכז העומד מאחוריו שימוש בטכניקת Dropout במהלך אימון רשתות נירוניים?
- א. Dropout מאייך את זמן האימון על ידי השמטת שכבות שלמות מהרשת בכלל אפוק.
  - ב. Dropout מסייע במניעת Overfitting על ידי הפעלת רשתות שונות בכל איטרציה, ובכך מדמה אנסמבל של מודלים.
  - ג. Dropout מפחית את עומק הרשת באופן זמני, וכך משפר את קצב ההתקנסות.
  - ד. Dropout משמש להקטנת ערכי הגרדיאנט בעת Backpropagation לצורך ייצוב האימון.

### שאלה 17

- במקרים מסוימים, הוספת רגולרייזציה מסווג  $L^2$  לאימון רשת הובילה להחמרה ביצועים על קבוצת הווילדי. מה מהבאים הוא הסבר אפשרי לכך?
- א. רגולרייזציה  $L^2$  תמיד משפרת את ביצועי המודל על קבוצת הווילדי, ולכן מדובר כנראה בטעות מדידה.
  - ב. הרגולרייזציה הגבוהה יוצרה מודל בעל פרמטרים שמרניים מדי, שפגעו ביכולת המודל לקלוט מבנים מורכבים - תופעה שמצויה עם Underfitting.
  - ג. רגולרייזציה  $L^2$  הופכת את פונקציית ה- Loss ללא גזירה, ולכן האימון אינו מתכנס כראוי.
  - ד. בגלל רגולרייזציה מסווג  $L^2$  המודל מעונייש פרמטרים גדולים וכאן מדובר בתכונות דומיננטיות בקלט, דבר שפוגע ביכולת ההכללה.

### שאלה 18

- כיצד משפיע שיטת ה- Dropout על התלות הגדזית (coupling) בין נירונים באותו שכבה בראשת נירונים? **בחרו את האפשרות המדויקת ביותר.**
- א. Dropout מגביר את התלות הגדזית (coupling) בין נירונים, משום שהוא מאלץ אותם ללמידה לעבוד יחד.
  - ב. Dropout מפחית את התלות הגדזית בין נירונים, בכך שהוא מאלץ כל ניירון ללמידה תכונות באופן עצמאי ולא להסתמך על נירונים אחרים.
  - ג. Dropout אינו משפיע על ה- coupling בין נירונים, אלא רק על קצב הלמידה וההתקנסות של המודל.
  - ד. Dropout גורם לדילול (sparsity) בקשרים, אך אינו משנה את אופן שיתוף הפעולה בין הנירונים.
  - ה. תשובה א', ב', ג' ו-ד' אינה נכונה.

### שאלה 19

ברשת Fully Connected Feedforward השבנה נעה שימוש בפונקציית האקטיבציה ReLU בכל השכבות החבויות, אתחול לא תקין של המשקלים עלול לגרום לבעה חמורה באימון. מהי ההשפעה השכיחה ביותר של בעיה זו? בחרו את האפשרות המדויקת ביותר.

- א. הרשות תיכשל בלמידת ייצוגים מורכבים בגל היעדר אי-ליניאריות.
- ב. חלק מהנוירונים יפסיקו להגיב (dead neurons) כבר בתחילת האימון, ולא יתרמו עוד ללמידה.
- ג. קצב הלמידה יוגבר מאוד, מה שעולל לגרום לחוסר יציבות.
- ד. תשובה א', ב', ו- ג' אינן נכונות.

### שאלה 20

נתונה רשת Fully Connected Feedforward הכוללת מספר שכבות חבויות, כאשר בכל שכבה כלנוירון מחובר לכלנוירון בשכבה הקודמת וככל גם איבר הטיה (bias). אחת השכבות החבויות כוללת  $M^2$ נוירונים. מוצעת רשת חלופית שבה מחליפים את אותה שכבה בשתי שכבות חבויות ברצף, כאשר כל שכבה כוללת  $M$ נוירונים, וכלנוירון בהן גם מחובר לכלנוירון בשכבה הקודמת וככל איבר הטיה.

איזה מהטענות הבאות נכונה ביותר לגבי מספר הפרמטרים הכולל ברשת המקורית לעומת רשת החלופית?

- א. מספר הפרמטרים ברשת המקורית **יותר** מאשר במספר הפרמטרים ברשת החלופית.
- ב. מספר הפרמטרים ברשת המקורית **קטן יותר** מאשר במספר הפרמטרים ברשת החלופית.
- ג. מספר הפרמטרים ברשת המקורית שווה במספר הפרמטרים הכולל ברשת החלופית.
- ד. לא ניתן לקבוע את היחס בין מספר הפרמטרים הכולל בשתי הרשומות (האם גדול, קטן או שווה) מבלי לדעת את מספר השכבות ברשת ואת מספר הנוירונים בכל שכבה.
- ה. תשובה א', ב', ג' ו-ד' אינן נכונות.

**ב ה צ ל ח ה !**

ריכוז הגדירות 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:

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

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

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