



מספר תעודת זהות
יש לרשום את כל תשע הספרות

האוניברסיטה
הפתוחה



יש להדביק כאן את
מדבקת הנבחן

כ"ז באב תשפ"ה

מס' שאלון - 572
21
באוגוסט 2025

סמסטר 2025ב

מס' מועד 93

22961 / 4

שאלון בחינת גמר
22961 - למידה עמוקה

משך בחינה: 3 שעות

בשאלון זה 11 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 כרטיס המחשב שמצורף 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

איזו מהשיטות הבאות נחשבת לשיטת רגולריזציה (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' אינן נכונות.

שאלה 2

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

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

שאלה 3

בהנחה שיש ברשותכם רשת מאומנת מראש על מיליוני תמונות של אובייקטים כלליים (כגון ImageNet), ואתם מעוניינים לאמן מודל חדש לזיהוי פתולוגיות רפואיות בצילומי רנטגן, אך במשימה החדשה יש מספר מצומצם של דוגמאות אימון, איזו אסטרטגיית Transfer Learning תיחשב לאפקטיבית ביותר? **בחרו את האפשרות המדויקת ביותר.**

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

שאלה 4

ברשתות קונבולוציה (CNN), נעשה שימוש בשכבות Pooling כחלק מהארכיטקטורה 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 ללמוד ייצוגים מורכבים יותר.
- ג. מאיצות את תהליך האופטימיזציה באמצעות שינוי הגרדיאנטים בשכבות הקונבולוציה.
- ד. מצמצמות את המימדים המרחביים של הנתונים, מפחיתות את מספר הפרמטרים, ומאפשרות למודל לזהות תכונות באופן חסין לשינויים במקום (translation invariance).
- ה. משפרות את דיוק המודל על ידי הקטנת שגיאת הניבוי באופן ישיר.

שאלה 5

במהלך יישום של אלגוריתם מורד הגרדיאנט (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 עשויה להיות תלויה בנקודת ההתחלה.
- ג. קצב הלמידה (learning rate) משפיע רק על מהירות ההתכנסות, אך אינו משפיע על איכות הפתרון הסופי.
- ד. כל עוד מתבצעים עדכונים לפי הגרדיאנט, מובטחת התכנסות לפתרון יציב.
- ה. אלגוריתם מורד הגרדיאנט 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 מסוג Fully Connected Feedforward, הכוללת שכבת קלט, שכבה חבויה אחת, ושכבת פלט אחת עם נוירון בודד לצורך סיווג בינארי. בכל שכבה, כל נוירון מקבל קלט מכל הנוירונים בשכבה הקודמת וכולל גם הטיה (bias). מימד הקלט הוא 20×20 . ידוע כי מספר הפרמטרים הכולל (כולל משקלים והטיות) ברשת זו הוא 5,629. מהו מספר הנוירונים בשכבה החבויה?

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

שאלה 7

במודלים של למידה מפקחת (supervised learning), מקובל להשתמש בגולריזציה מסוג L^1 או גולריזציה מסוג L^2 כדי להפחית סיכון להתאמת יתר (overfitting). מהו ההבדל המרכזי בין שתי שיטות הרגולריזציה? **בחרו את האפשרות המדויקת ביותר.**

- א. רגולריזציה מסוג L^2 נוטה לאפס חלק מהמשקלים לחלוטין, ולכן יוצרת פתרונות דלילים (sparse).
- ב. רגולריזציה מסוג L^1 שומרת על כל הפרמטרים קטנים אך לא מאפסת אותם.
- ג. רגולריזציה מסוג L^1 משמרת את כל הפרמטרים בתוצאה, ולכן אינה מתאימה לצמצום מספר הפיצ'רים הפעילים במודל.
- ד. רגולריזציה מסוג L^1 תמיד עדיפה על רגולריזציה מסוג L^2 , במיוחד ברשתות נוירונים עמוקות.
- ה. תשובות א', ב', ג' ו-ד' נכונות.
- ו. תשובות א', ב', ג' ו-ד' אינן נכונות.

שאלה 8

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

- א. להפסיק את האימון כאשר השיפור בביצועי המודל על קבוצת האימון נעצר או מתחיל להיפגע, כדי לחסוך בזמן חישוב.
- ב. להפסיק את האימון כאשר הביצועים על קבוצת הוולידציה מפסיקים להשתפר ומתחילים להידרדר, במטרה למנוע התאמת 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 לסף מסוים ומתייצבת.
- ה. להפסיק את האימון כשפונקציית ה-Loss משתפרת בהתמדה לאורך מספר אפוקים רצופים.
- ו. תשובות א', ב', ג', ד' ו-ה אינן נכונות.

שאלה 9

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

- א. לנרמל או לקודד את נתוני הקלט לפני המעבר לשכבות הבאות.
- ב. לאפשר לרשת ללמוד קשרים לא ליניאריים בין הקלט לפלט.
- ג. לאפשר למודל להתכנס במהירות במהלך תהליך האימון.
- ד. להפחית את הסיכון להתאמת יתר (overfitting) באמצעות רגולריזציה פנימית.
- ה. תשובות א', ב', ג' ו-ד אינן נכונות.

שאלה 10

במהלך אימון רשתות נוירונים עמוקות במיוחד (למשל, בעלות כ-100 שכבות) למשימות סיווג בינארי, נעשה שימוש בפונקציית Sigmoid בשכבת הפלט בשילוב של פונקציות אקטיבציה מסוג tanh ו-ReLU בשכבות הפנימיות. לאחר epoch אחד בלבד, מתברר כי בשכבות מסוימות המשקלים מפסיקים להתעדכן, אף שהרשת לא התכנסה. בדיקה מעמיקה מעלה כי הגרדיאנטים באותן שכבות נעלמים כמעט לגמרי כבר בשלבים מוקדמים של האימון, בעוד שפונקציית ההפסד (Loss) נשארת בטווח סביר.

בהתחשב בכך, אילו מהשינויים האפשריים הבאים עשויים לסייע בצמצום הבעיה?

1. להגדיל את גודל קבוצת האימון.
2. להחליף את פונקציות ה-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 שינוי אחד מתוך הארבעה עשוי לסייע בצמצום הבעיה.
- ב. רק שני שינויים מתוך הארבעה עשויים לסייע בצמצום הבעיה.
- ג. שלושה שינויים מתוך הארבעה עשויים לסייע בצמצום הבעיה.
- ד. כל ארבעת השינויים המוצעים עשויים לסייע בצמצום הבעיה.
- ה. אף אחד מהשינויים המוצעים אינו עשוי לסייע בצמצום הבעיה.

שאלה 11

Dropout היא טכניקת רגולריזציה נפוצה באימון רשתות נוירונים.

בחרו בתשובה המדויקת ביותר.

- א. Dropout מוביל לדלילות (sparsity) במשקלים המאומנים של הרשת.
- ב. במהלך שלב הבדיקה (test time), Dropout מיושם עם הסתברות הישארות (keep probability) הפוכה.
- ג. ככל שהסתברות ההישארות גבוהה יותר, כך האפקט של הרגולריזציה 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) ריבועית עם איבר רגולריזציה מסוג Ridge 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

$$\mathbf{v}_1^T = [1, 2]/\sqrt{5}, \mathbf{v}_2^T = [-2, 1]/\sqrt{5}$$

עצמיים מתאימים $\mathbf{v}_1^T = [1, 2]/\sqrt{5}, \mathbf{v}_2^T = [-2, 1]/\sqrt{5}$ ידוע כי עבור ערך רגולריזציה $\lambda = 10$, פתרון בעיית המינימום הוא

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

שאלה 12

מהו הפתרון \mathbf{w}^* הממוזער את פונקציית העלות כאשר אין איבר רגולריזציה (כלומר, כאשר

$$\lambda = 0)?$$

א. $\mathbf{w}^{*T} = [1/2, -1/2]$

ב. $\mathbf{w}^{*T} = [-1/2, 1/2]$

ג. $\mathbf{w}^{*T} = [1/2, 0]$

ד. $\mathbf{w}^{*T} = [0, -1/2]$

ה. $\mathbf{w}^{*T} = [1/2, -1/25]$

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

שאלה 13

הוקטור \mathbf{b} נתון ע"י

א. $\mathbf{b} = [-0.5, 2.75]^T$

ב. $\mathbf{b} = [1, -5.5]^T$

ג. $\mathbf{b} = [-1, 5.5]^T$

ד. $\mathbf{b} = [0.5, -2.75]^T$

ה. $\mathbf{b} = [-0.25, 1.375]^T$

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

שאלה 14

רשת קונבולוציה מקבלת כקלט טנזור ממימד 256×64 העובר דרך שכבת קונבולוציה בעלת 32 פילטרים עם גודל גרעין 10×6 , 5×3 stride, וללא ריפוד באפסים. מה יהיה גודל טנזור 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 Stochastic Gradient Descent (SGD), נדרש לבחור גודל מתאים למיני-באץ' (mini-batch). מהם היתרונות והחסרונות של שימוש במיני-באץ' קטנים לעומת מיני-באץ' גדולים?

בחרו את התשובה הנכונה ביותר.

- א. מיני-באץ' קטנים מובילים לעדכוני משקל תכופים יותר, עם גרדיאנטים רועשים יותר, מה שיכול לסייע ביציאה ממנימום מקומי אך עשוי להאט את ההתכנסות ולדרוש קצב למידה קטן יותר.
- ב. מיני-באץ' גדולים מפחיתים את רעש הגרדיאנט, מה שתורם להתכנסות יציבה יותר במספר צעדים קטן יותר, אך הם דורשים יותר זיכרון חישובי ועלולים להיתקע במנימום מקומי.
- ג. מיני-באץ' גדולים צורכים פחות משאבי זיכרון ולכן הם תמיד עדיפים חישובית.
- ד. בפועל, גודל המיני-באץ' נבחר לעתים קרובות על פי מגבלות זכרון ה-GPU ולא רק משיקולי אופטימיזציה תיאורטיים.
- ה. שימוש מיני-באץ' קטנים מגביר את הסיכון להתאמת יתר (overfitting) ולכן יש להימנע מהם.
- ו. בין התשובות א', ב', ג', ד' ו-ה יש יותר מתשובה אחת נכונה.

שאלה 16

- מהו העיקרון המרכזי העומד מאחורי שימוש בטכניקת Dropout במהלך אימון רשתות נוירונים?
- א. Dropout מאיץ את זמן האימון על ידי השמטת שכבות שלמות מהרשת בכל אפוק.
- ב. Dropout מסייע במניעת Overfitting על ידי הפעלת רשתות משנה שונות בכל איטרציה, ובכך מדמה אנסמבל של מודלים.
- ג. Dropout מפחית את עומק הרשת באופן זמני, וכך משפר את קצב ההתכנסות.
- ד. Dropout משמש להקטנת ערכי הגרדיאנט בעת Backpropagation לצורך ייצוב האימון.

שאלה 17

- במקרים מסוימים, הוספת רגולריזציה מסוג L^2 לאימון רשת הובילה להחמרה בביצועים על קבוצת הוולידציה. מה מהבאים הוא הסבר אפשרי לכך?
- א. רגולריזציית L^2 תמיד משפרת את ביצועי המודל על קבוצת הוולידציה, ולכן מדובר כנראה בטעות מדידה.
- ב. הרגולריזציה הגבוהה יצרה מודל בעל פרמטרים שמרניים מדי, שפגעו ביכולת המודל לקלוט מבנים מורכבים - תופעה שמזוהה עם Underfitting.
- ג. רגולריזציית L^2 הופכת את פונקציית ה-Loss ללא גזירה, ולכן האימון אינו מתכנס כראוי.
- ד. בגלל רגולריזציה מסוג L^2 המודל מעניש פרמטרים גדולים ולכן מדכא תכונות דומיננטיות בקלט, דבר שפוגע ביכולת ההכללה.

שאלה 18

- כיצד משפיעה שיטת ה-Dropout על התלות הדדית (coupling) בין נוירונים באותה שכבה ברשת נוירונים? **בחרו את האפשרות המדויקת ביותר.**
- א. Dropout מגביר את התלות ההדדית (coupling) בין נוירונים, משום שהוא מאלץ אותם ללמוד לעבוד יחד.
- ב. Dropout מפחית את התלות ההדדית בין נוירונים, בכך שהוא מאלץ כל נוירון ללמוד תכונות באופן עצמאי ולא להסתמך על נוירונים אחרים.
- ג. Dropout אינו משפיע על ה-coupling בין נוירונים, אלא רק על קצב הלמידה וההתכנסות של המודל.
- ד. Dropout גורם לדילול (sparsity) בקשרים, אך אינו משנה את אופן שיתוף הפעולה בין הנוירונים.
- ה. תשובות א', ב', ג' ו-ד' אינן נכונות.

שאלה 19

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

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

שאלה 20

נתונה רשת Fully Connected Feedforward הכוללת מספר שכבות חביויות, כאשר בכל שכבה כל נוירון מחובר לכל נוירון בשכבה הקודמת וכולל גם איבר הטיה (bias). אחת השכבות החביויות כוללת $2M$ נוירונים. מוצעת רשת חלופית שבה מחליפים את אותה שכבה בשתי שכבות חביויות ברצף, כאשר כל שכבה כוללת M נוירונים, וכל נוירון בהן גם מחובר לכל נוירון בשכבה הקודמת וכולל איבר הטיה.

איזו מהטענות הבאות נכונה ביותר לגבי מספר הפרמטרים הכולל ברשת המקורית לעומת הרשת החלופית?

- א. מספר הפרמטרים ברשת המקורית **גדול יותר** ממספר הפרמטרים ברשת החלופית.
- ב. מספר הפרמטרים ברשת המקורית **קטן יותר** ממספר הפרמטרים ברשת החלופית.
- ג. מספר הפרמטרים ברשת המקורית **שווה** למספר הפרמטרים הכולל ברשת החלופית.
- ד. לא ניתן לקבוע את היחס בין מספר הפרמטרים הכולל בשתי הרשתות (האם גדול, קטן או שווה) מבלי לדעת את מספר השכבות ברשת ואת מספר הנוירונים בכל שכבה.
- 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 $\mathbf{x}^{(0)}$:

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