



מספר תעודת זהות
יש לרשום את כל תשע הספרות

האוניברסיטה
הפתוחה



יש להדביק כאן את
מדבקת הנבחן

י"ד בתמוז תשפ"ה

מס' שאלון - 572
10 ביולי 2025

סמסטר 2025ב

22961 / 4

מס' מועד 86

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

איזו מהתשובות הבאות נכונה ביותר עבור פעולת אלגוריתם מורד הגרדיאנט (gradient descent) לתהליך המינימיזציה של פונקציית המחיר (cost function).

- א. בכל נקודה במרחב הפרמטרים, תנועה קטנה בכיוון השלילי של הגרדיאנט תוביל לירידה בערך פונקציית המחיר.
- ב. אם גודל הצעד (learning rate) קטן מדי, האלגוריתם עשוי להתקדם באיטיות רבה ואף לא להתכנס גם לאחר מספר רב של איטרציות.
- ג. גודל צעד גדול מדי יוביל לאי-יציבות בתהליך וייתכן שנדלג מעל נקודת המינימום, וכך ערך פונקציית המחיר לא יירד.
- ד. כל התשובות נכונות.

שאלה 2

נתונות הטענות הבאות עבור פונקציית ה-softmax:

- 1. מניבה מספרים חיוביים שסכומם אחת.
- 2. אינוואריאנטית לתוספת קבוע לכל אחד מהאיברים, כלומר $\text{softmax}(x_1 + c, x_2 + c, \dots, x_n + c) = \text{softmax}(x_1, x_2, \dots, x_n)$.
- 3. פונקציה גזירה.
- 4. תמיד מחזירה ערכים בינאריים (0 או 1).
- 5. אינה מושפעת מסדר האיברים בוקטור הקלט.

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. רק ארבע טענות נכונות.
- ו. כל חמש הטענות נכונות.

שאלה 3

השימוש בשיטת הגרדיאנט האקראי (Stochastic Gradient Descent - SGD) נועד להתמודד עם הבעיה הבאה (בחרו בתשובה הנכונה ביותר):

- א. היעדר גרדיאנט בנקודות לא גזירות של פונקציית העלות.
- ב. חישוב איטי של הגרדיאנט באימון רשתות מורכבות על אוסף נתונים (dataset) גדול.
- ג. חוסר יכולת למצוא מינימום גלובלי של פונקציית העלות.
- ד. תשובות א', ב', ו-ג' נכונות.
- ה. תשובות א', ב', ו-ג' אינן נכונות.

שאלה 4

נתונה רשת נוירונים מסוג Fully connected FeedForward הכוללת שתי שכבות, כאשר פלט הרשת מוגדר ע"י:

$$y = h(U^T g(W^T x + b) + c)$$

עם פונקציות אקטיבציה $g()$ בשכבה החבויה ו- $h()$ בשכבת המוצא. נתון כי וקטור הקלט הוא $x \in \mathbb{R}^d$, הפרמטרים הנלמדים הם מטריצות המשקלות $W \in \mathbb{R}^{d \times k}$, $U \in \mathbb{R}^{k \times d}$, ווקטורי ההטיות $b \in \mathbb{R}^k$ ו- $c \in \mathbb{R}^d$, והפלט הוא $y \in \mathbb{R}^d$. נתון גם ש- $k < d$.

בחרו בתשובה הנכונה ביותר, בהינתן שפונקציות האקטיבציה הן פונקציות הזהות (identity), כלומר:

$$g(z) = z, h(z) = z$$

- רשת זו מסוגלת לייצר כל טרנספורמציה אפינית (affine) מהצורה $y = P^T x + d$ בין הקלט $x \in \mathbb{R}^d$ לבין הפלט $y \in \mathbb{R}^d$.
- רשת זו מאפשרת ייצוג של הקלט $x \in \mathbb{R}^d$ באמצעות מרחב נמוך מימד.
- הודות לשימוש ביותר משכבה אחת, רשת זו עשויה לייצג קשרים מורכבים יותר מקשרים אפיניים (affine) בין הקלט לפלט.
- תשובות א', ב', ו-ג' אינן נכונות.
- תשובות א', ב', ו-ג' נכונות.

שאלה 5

מה ההבדל העיקרי בין Stochastic Gradient Descent (SGD) עם מומנטום לבין Adam ?

- SGD עם מומנטום משתמש רק במידע על כיוון הגרדיאנט, בעוד Adam מתאים את כיוון וגודל העדכון לפי גרדיאנטים קודמים.
- Adam משלב מומנטום עם התאמה של קצב הלמידה לכל פרמטר בנפרד.
- בעוד ש-SGD עם מומנטום משתמש בגרדיאנט ממוצע לאורך זמן, Adam מחשב עדכון מבוסס על סטיית תקן של הגרדיאנטים בלבד.
- Adam משתמש רק בגרדיאנט הנוכחי, ללא מידע היסטורי כמו מומנטום.

שאלה 6

מהם היתרונות המרכזיים של רשתות עצביות קונבולוציוניות (Convolutional Neural Networks) בעיבוד תמונות, בהשוואה ל-Fully connected FeedForward networks?

- רשתות קונבולוציה דורשות יותר פרמטרים ולכן מסוגלות ללמוד תבניות מורכבות יותר.
- רשתות קונבולוציה משתמשות בשכבות קונבולוציה המאפשרות שיתוף משקלים וזיהוי תבניות מקומיות, דבר המפחית את מספר הפרמטרים ומשפר את יכולת ההכללה.
- רשתות קונבולוציה אינן תלויות בסדר הפיקסלים ולכן אינן מושפעות ממיקום האובייקטים בתמונה.
- רשתות קונבולוציה מתעלמות ממבנה התמונה ולכן אינן מנצלות מידע מרחבי לצורך זיהוי תכונות.
- תשובות א', ב', ג' ו-ד' נכונות.
- תשובות א', ב', ג' ו-ד' אינן נכונות.

שאלה 7

במסגרת אימון רשת נוירונים לזיהוי עצמים בתמונות, המודל מפגין דיוק גבוה על קבוצת האימון אך ביצועים ירודים על קבוצת הוולידציה. איזו מהאפשרויות הבאות פחות מתאימה להתמודדות עם מצב זה?

- עצירה אוטומטית של תהליך האימון כאשר אין שיפור מתמשך במדדי הביצוע על סט הוולידציה.
- השמטה אקראית של יחידות פנימיות בשכבות הנסתרות במהלך האימון.
- שילוב של פלטים ממספר עותקים שונים של הרשת שאומנו בנפרד.
- הגדלת מספר שכבות הרשת והארכת זמן האימון עד למיצוי מוחלט של הדאטה.
- יצירת גרסאות נוספות של הדאטה באמצעות סיבובים, חיתוכים ושינויים חזותיים.

שאלה 8

רשת קונבולוציה מקבלת כקלט טנזור קלט ממיד 128×256 העובר דרך שכבת קונבולוציה בעלת 64 פילטרים עם גודל גרעין 10×6 , 3×5 stride, וללא ריפוד באפסים. מה יהיה גודל טנזור הפלט?

בחרו את התשובה הנכונה:

- $64 \times 40 \times 50$.
- $64 \times 40 \times 51$.
- $64 \times 39 \times 50$.
- $64 \times 41 \times 52$.
- תשובות א', ב', ג' ו-ד' אינן נכונות.

שאלה 9

צוות חוקרים מבקש לאמן מודל לזיהוי סוגים נדירים של סרטן באמצעות תמונות מיקרוסקופיות. ברשותם רק 300 תמונות מתויגות, ומספר הקטגוריות לסיווג גבוה יחסית.

באיזו מהגישות הבאות סביר ביותר להשתמש כדי לשפר את ביצועי המודל בתנאים אלו?

- לאמן רשת עמוקה מאוד מאפס (from scratch) על סט התמונות הקטן.
- להשתמש ב- Transfer Learning עם מודל שאומן מראש על תמונות כלליות (כגון ImageNet) ולבצע fine-tuning.
- לשכפל את התמונות הקיימות כדי להרחיב את סט האימון.
- לאמן רשת עמוקה מאוד מאפס (from scratch), ולהחיל Dropout על כל שכבות המודל כדי לצמצם תופעת אימון היתר (overfitting).

שאלה 10

איזו מהדרכים הבאות אינה נחשבת לשיטה מקובלת ליישום Transfer Learning במודל רשת עמוקה?

- הקפאת (freeze) השכבות המוקדמות במודל שאומן מראש ואימון רק של השכבות האחרונות על הדאטה החדש.
- שכפול ארכיטקטורת המודל המאומן מראש, אך אתחול אקראי של כל המשקולות ואימון מחדש על הדאטה החדש כדי למנוע הטיות ממשימה קודמת.
- טעינת משקולות ממודל שאומן מראש וביצוע fine-tuning על חלק מהשכבות לפי הדאטה החדש.
- החלפת השכבות האחרונות במודל המאומן בשכבות חדשות המותאמות למשימה חדשה.

שאלה 11

בעת תכנון רשת נוירונים, קיימת התאמה טבעית בין סוג הבעיה, פונקציית האקטיבציה בשכבת הפלט, ופונקציית המחיר (cost function). איזו מההתאמות הבאות אינה אידיאלית לסוג הבעיה?

- בעיית רגרסיה, פונקציית פלט: זהות (identity), פונקציית המחיר: סכום ריבועים.
- סיווג בינארי, פונקציית פלט: סיגמואיד, פונקציית המחיר: Cross Entropy.
- סיווג רב-מחלקתי (multi-class), פונקציית פלט: softmax, פונקציית המחיר: Cross Entropy רב-מחלקתי.
- סיווג בינארי, פונקציית פלט: tanh, פונקציית המחיר: סכום ריבועים.

שאלה 12

נתונה פונקציה ריבועית קמורה מהצורה:

$$f(x) = \frac{1}{2} x^T A x$$

כאשר A היא מטריצה סימטרית וחיובית מוגדרת. נניח כי קיים פער משמעותי בין הערך העצמי הקטן ביותר לערך העצמי הגדול ביותר של A .

אופטימיזציה של הפונקציה מתבצעת באמצעות שיטת מורד הגרדיאנט (Gradient Descent) עם וקטור התחלה שרירותי ועם גודל צעד קבוע ותקין (שאינו גדול מדי כך שהשיטה מתכנסת תאורטית). הסבירו מהו הגורם העיקרי שמשפיע על קצב ההתכנסות של השיטה, ובחרו בתשובה הנכונה ביותר מתוך האפשרויות הבאות:

- א. ההתכנסות מהירה בגלל שהגרדיאנטים גדולים בכיוון של הערך העצמי הגדול של A .
- ב. ההתכנסות איטית, בעיקר בגלל שהעדכוני בכיוון של הערך העצמי הגדול של A גדולים יותר.
- ג. ההתכנסות איטית, בעיקר בגלל שהעדכוני בכיוון של הערך העצמי הקטן של A קטנים יותר.
- ד. שיטת מורד הגרדיאנט אינה רגישה כלל לערכים העצמיים של A .
- ה. תשובות א', ב', ג' ו-ה אינן נכונות.

שאלה 13

מהי מטרת השימוש ב־ Dataset Augmentation באימון רשתות נוירונים?

בחרו איזו מהתשובות הבאות אינה נכונה.

- א. להרחיב את גודל קבוצת האימון על ידי יצירת דוגמאות נוספות מהדאטה הקיים.
- ב. לשפר את יכולת ההכללה של המודל ולצמצם סכנת Overfitting.
- ג. לקצר את זמן האימון הכולל של המודל, מאחר והמודל רוכש את ההתפלגות הרצויה של הנתונים תוך שימוש בפחות דוגמאות אמיתיות.
- ד. לאלץ את המודל ללמוד אינוואריאנטיות מסוימות שהתופעה אמורה להיות אדישה להן.

שאלה 14

מה ההשפעה הצפויה של dropout על זמן האימון הנדרש להתכנסות? **בחרו בתשובה הנכונה ביותר.**

- א. מקצר את זמן האימון מאחר ופחות נוירונים פעילים.
- ב. מאריך את זמן האימון, כי בכל איטרציה מאמנים תת-רשת שונה.
- ג. אין כל השפעה על זמן האימון.
- ד. Dropout מפסיק את האימון ברגע שנוצר overfitting, ולכן מקצר את זמן האימון.
- ה. תשובות א', ב', ג' ו-ד אינן נכונות.

שאלות 15,16:

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 $\lambda_1 = 10, \lambda_2 = 2.5$ ווקטורים עצמיים מתאימים $v_1^T = [1, 2]/\sqrt{5}, v_2^T = [-2, 1]/\sqrt{5}$. הוקטור \mathbf{b} נתון ע"י $\mathbf{b}^T = [-1, 5.5]$.

שאלה 15

חשבו את הנקודה \mathbf{w}^* המביאה למינימום את פונקציית המחיר.

א. $\mathbf{w}^{*T} = [1, -1]$

ב. $\mathbf{w}^{*T} = [-1, 1]$

ג. $\mathbf{w}^{*T} = [1, 0]$

ד. $\mathbf{w}^{*T} = [0, -1]$

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

שאלה 16

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. $\mathbf{w}_{reg}^T \approx [-0.14, 0.32]$

ב. $\mathbf{w}_{reg}^T \approx [-0.14, 0.14]$

ג. $\mathbf{w}_{reg}^T \approx [0.14, -0.32]$

ד. $\mathbf{w}_{reg}^T \approx [0.32, -0.14]$

ה. $\mathbf{w}_{reg}^T \approx [-0.32, 0.14]$

ו. $\mathbf{w}_{reg}^T \approx [0.14, -0.14]$

שאלה 17

מהו היתרון העיקרי של פונקציית האקטיבציה ReLU בהשוואה ל- sigmoid ול- tanh ברשתות נוירונים? **בחרו בתשובה הנכונה ביותר.**

- א. פונקציית ה-ReLU מבטיחה שפלט האקטיבציה יהיה בין 0 ל- 1, ולכן מתאימה טוב יותר לבעיות סיווג.
- ב. פונקציית ה-ReLU אינה גורמת לעולם לבעיה של גרדיאנט נעלם (vanishing gradients).
- ג. פונקציית ה-ReLU חוסכת חישובים ונוטה להוביל להתכנסות מהירה יותר באימון רשתות נוירונים.
- ד. פונקציית ה-ReLU משמרת מידע מרחבי טוב יותר.

שאלה 18

במהלך אימון רשת CNN הכוללת מספר שכבות convolution ו-MaxPooling, אתם מבחינים בכך שמידע מרחבי חשוב הולך לאיבוד כבר בשכבות הראשונות. **מה הסיבה הסבירה ביותר לכך?**

- א. ערך ה-stride בשכבות הקונבולוציה גדול מדי.
- ב. פונקציית האקטיבציה ReLU גורמת לאובדן מידע מרחבי ולכן אינה מתאימה למבנה כזה.
- ג. החסר בשכבת Fully Connected גורם לכך שהרשת לא מצליחה לשמר מידע בשכבות הראשונות.
- ד. אין מספיק training data, ולכן הרשת לא לומדת לשמר את המידע המרחבי החשוב.

שאלה 19

בעת חישוב הגרדיאנט של פונקציית המחיר (cost function) ביחס לפרמטרים של שכבה מסוימת ברשת נוירונים, אילו רכיבים משפיעים עליו? **בחרו בתשובה הנכונה ביותר:**

- א. רק הקלט שהוזן לשכבה זו.
- ב. רק השגיאה שחושבה ביציאת הרשת.
- ג. גם הקלט לשכבה וגם הגרדיאנט שהועבר מהשכבה הבאה.
- ד. רק גודל הרשת.
- ה. תשובות א', ב', ג', ו-ד' אינן נכונות.

שאלה 20

נתונה רשת נוירונים לסיווג בינארי עם שכבה אחת, שבה הפלט מחושב לפי:

$$\hat{y} = \sigma(\mathbf{w}^T \mathbf{x} + b)$$

כאשר σ היא פונקציית ה-sigmoid, ופונקציית המחיר J היא cross-entropy, כלומר עבור דוגמה בודדת (\mathbf{x}, y) :

$$J(\mathbf{x}, y) = -(y \log(\hat{y}) + (1 - y) \log(1 - \hat{y}))$$

מהו הגרדיאנט של פונקציית המחיר $J(\mathbf{x}, y)$ ביחס לוקטור הפרמטרים \mathbf{w} ?

א. $(2y - 1)\sigma((1 - 2y)(\mathbf{w}^T \mathbf{x} + b))\mathbf{x}$

ב. $(1 - 2y)\sigma((1 - 2y)(\mathbf{w}^T \mathbf{x} + b))\mathbf{x}$

ג. $(2y - 1)\left(1 - \sigma((1 - 2y)(\mathbf{w}^T \mathbf{x} + b))\right)\mathbf{x}$

ד. $(2y - 1)\sigma((2y - 1)(\mathbf{w}^T \mathbf{x} + b))\mathbf{x}$

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

בהצלחה!

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

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

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

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