

שאלות חזרה – למידה חישובית

1. למה המחיר לא כלול בתכונות?
 - המחיר הוא המשתנה המוסבר (dependent variable), כלומר המטרה שלנו היא לחזות אותו על סמך התכונות האחרות (כגון שטח, קומה וכו'). לכן, הוא לא נכלל בתוך מטריצת הקלט X , אלא נשמר כוקטור מטרה y .
2. למה אנחנו מבצעים תקנון (נרמול) לתכונות לפני הלמידה עם ADAGRAD?
 - כדי שכל תכונה תהיה באותו סדר גודל ולא תשלוט על הגרדיאנט. זה חשוב במיוחד באלגוריתמים כמו ADAGRAD שמבצעים עדכון לכל פרמטר לפי גודלו.
3. למה מוסיפים bias למטריצת הקלט?
 - ה-bias מאפשר למודל לקבל הזזה קבועה ולשפר התאמה גם כאשר כל הקלטים הם אפס. זה כמו לאפשר למודל "נקודת התחלה" שלא תלויה בתכונות.
4. מה ההבדל בין הפתרון הנורמלי ל-ADAGRAD?
 - הפתרון הנורמלי הוא **פתרון אנליטי** – משתמש במטריצות לפתרון ישיר.
 - ADAGRAD הוא אלגוריתם איטרטיבי – לומד שלב-שלב ומעדכן פרמטרים לפי הגרדיאנט.
5. למה נרמלנו גם את המחיר (y) לפני השימוש ב-ADAGRAD?
 - כדי שהסקלה של y לא תגרום לבעיות בלמידה (למשל מחירים מאוד גבוהים שגורמים לעלות גבוהה מדי), וזה מקל על התכנסות המודל.
6. מה המשמעות של cost ולמה עוקבים אחריו בכל איטרציה?
 - ה-cost (למשל MSE) מודד כמה התחזיות של המודל רחוקות מהתוצאות האמיתיות. עוקבים אחריו כדי לראות אם המודל משתפר.
7. למה צריך את וקטור θ (טטא)?
 - וקטור θ (טטא) מייצג את הפרמטרים של המודל – כלומר, המשקלים שמראים כמה כל תכונה משפיעה על התחזית של המחיר.
 - בכל איטרציה של למידה (או בפתרון הנורמלי), אנחנו מעדכנים את טטא כדי שהמודל יתאים טוב יותר לנתונים.
8. מה זה נרמול / תקנון (Normalization/Standardization) ולמה צריך את זה?
 - נרמול (תקנון) פירושו להמיר את כל התכונות לסקלה אחידה, בדרך כלל כך שהממוצע הוא 0 והסטיית תקן היא 1.
 - זה חשוב במיוחד באלגוריתמים כמו Gradient Descent / ADAGRAD, כי תכונות עם ערכים גדולים ישפיעו חזק יותר על הלמידה ויגרמו לאי-יציבות.

מטרות עיקריות:

- למנוע שתכונה אחת תשלוט על הלמידה בגלל גודל המספרים שלה.
 - לזרז את ההתכנסות של האלגוריתם.
 - לשפר את הדיוק של המודל.
9. **F-score** (או בשמו המלא: F1-Score) הוא מדד ביצוע חשוב מאוד שמאזן בין דיוק (Precision) ל-רגישות (Recall), במיוחד כשיש לך נתונים לא מאוזנים (למשל: הרבה יותר דירות זולות מאשר יקרות, או הרבה יותר שליליים מחיוביים).

F1-Score הוא ממוצע הרמוני בין Precision ל-Recall.
Precision (דיוק) – כמה מהתחזיות החיוביות שלך היו נכונות בפועל.
Recall (או TPR / Sensitivity) – כמה מהמקרים החיוביים באמת הצלחת לזהות.

10. מה זה אלגוריתם KNN?

- KNN הוא אלגוריתם למידת חיזוי לא מפוקחת ישירה שמבצע סיווג (או רגרסיה) על פי השכנים הקרובים ביותר לדוגמה החדשה.
- הוא לא לומד משוואה – הוא שומר את כל הנתונים בזיכרון.
- כדי לסווג נקודה חדשה – הוא בודק מי k השכנים הכי קרובים אליה, ורואה מה הסיווג שלהם.

11. מה זה גרף ROC ומה הוא מודד?

- גרף ROC מודד את הביצוע של מודל סיווג ע"י הצגת היחס בין TPR (רגישות) ל-FPR (שיעור חיוביות שגויות).

12. מה המשמעות של כל נקודה בגרף?

- כל נקודה מייצגת סף (threshold) אחר, ולפי הסף רואים את ה-TPR וה-FPR של המודל.

13. למה צריך את כל המדדים (F1, Precision, FPR, TPR וכו')?

- כי אין מדד אחד שמספר את כל הסיפור:
- TPR אומר לנו כמה טוב המודל מזהה חיוביים.
- FPR אומר כמה הוא טועה עם שליליים.
- Precision אומר לנו כמה התחזיות החיוביות באמת נכונות.
- F1 מאזן בין Precision ל-Recall.

14. מה היתרון של ADAGRAD לעומת Gradient Descent רגיל?

- ADAGRAD מתאים את קצב הלמידה לכל פרמטר בנפרד.
- זה טוב כשיש תכונות בסקלה שונה או sparse data.
- לעומת זאת, GD משתמש בקצב למידה קבוע.

15. מה ההבדל בין ADAGRAD ל-Gradient Descent רגיל (GD)?

- Gradient Descent רגיל (GD) משתמש בקצב למידה קבוע (α) לכל הפרמטרים לאורך כל האיטרציות. כלומר, כל פרמטר במודל מתעדכן באותה עוצמה יחסית לגרדיאנט שלו, ללא קשר להיסטוריה של השינויים שלו.
- לעומת זאת, ADAGRAD הוא שיפור של GD, שבו קצב הלמידה מותאם באופן דינמי לכל פרמטר בנפרד, בהתאם להיסטוריה של הגרדיאנטים שלו.

ב-ADAGRAD:

- פרמטרים שקיבלו גרדיאנטים גדולים בעבר, יקבלו קצב למידה קטן יותר בעתיד.
- פרמטרים עם גרדיאנטים קטנים, יקבלו קצב למידה גבוה יותר.
- זה מאפשר למידה יציבה יותר, במיוחד בנתונים לא מאוזנים.

16. מה ההבדל בין רגרסיה לסיווג?

- רגרסיה מחזירה ערך רציף (מחיר),
סיווג מחזיר קטגוריה (למשל: יקרה או זולה).

