מבוא למערכות לומדות - חורף 22/23

236756

**HW4 dry**

**לוי הורביץ**

313511602

****

**שאלה 1**

**סעיף 1:**

כן, f קמורה.

**סעיף 2:**

נגדיר:

נשים לב שלכל X > 0 ו- X < 0 הפונ' גזירה ומתקיים ובנוסף מתקיים שגם גזירה (פנוקציות אלמנטריות).

נוכיח שהיא קמורה בתחום הנ"ל:

עבור  *מתקיים וכן* עבור  *מתקיים , לכן*  קמורה *לפי משפט.*

ומכיוון ש- גזירה וקמורה אז התת - גרדיאנט של הוא הגרדיאנט שלה עצמה (\*), ולכן מתקיים שלכל בתחום הנ"ל אכן *.*

נשאר להוכיח במקרה ש X = 0 שאכן *.*

על מנת להוכיח שאכן , צריך להוכיח על פי ההגדרה שמתקיים:

*נשים לב שעל פי הגדרת f, אכן מתקיים:*

*והוכחנו ש-. מש"ל.*

*(\**) לפי ההגדרה:

וזה מתקיים באופן מיידי מתכונת קמירות

**סעיף 3:**

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| *-2* | *1* | -1 | *0* |
| *-1* | *1/4* | *-0.5* | *1* |
| *-0.5* | *1/16* | *-0.25* | *2* |
| *-0.25* | *1/64* | *-1/8* | *3* |
| *-1/8* | *1/236* | *-1/16* | *4* |

*אפשר לראות לאחר 5 איטרציות שערכי X הולכים ומתקרבים ל- 0 (משמאל לציר Y) ובהתאמה ערכי הפונקציה הולכים וקרבים ל 0 עם מחזוריות שיטתית (ערכי הפונקציה נחתכים ברבע בכל איטרציה).*

**סעיף 4:**

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| *-2* | *1* | -1 | *0* |
| *2* | *2* | *1* | *1* |
| *-2* | *1* | *-1* | *2* |
| *2* | *2* | *1* | *3* |
| *-2* | *1* | *-1* | *4* |

קל לראות לאחר 5 איטרציות שהפונקציה תקועה על הערכים 1 ו 2 ולכן הפונקציה לא תתכנס לערך המינימלי שלה שהוא ב 0.

**שאלה 2**

*הסברים למעברים:*

1. *על פי כלל ההסתברות השלמה.*
2. *X הוא בלתי תלוי ב-w, ולכן ניתן להשמיט את ההתניה ב-w.*
3. *היא סקלר שאינו תלוי בארגומנט אותו אנו מחפשים שמביא את הביטוי למקסימום, ולכן ניתן להשמיט אותו.*
4. *Log היא פונקציה מונוטונית עולה, ולכן הפעלתו על הביטוי אינו משנה את הארגומנט שמביא את הביטוי למקסימום.*
5. *על פי חוקי לוגריתמים.*
6. *על פי הנתון: , ומכיוון ש- הוא הרעש של y בהינתן שאנו מסתכלים על x כלשהו, אז בעצם אנחנו מסתכלים רק על איך y יכול להתפלג כאשר x שלנו הוא קבוע. לכן בעצם אנחנו מקבלים: כי ההתפלגות של y נקבעת רק על פי שהוא מתפלג כפי שהסברנו לעיל.*
7. *על פי חוקי לוגריתמים.*
8. *על פי חוקי לוגריתמים.*
9. *הוא סקלר שאינו תלוי בארגומנט אותו אנו מחפשים שמביא את הביטוי למקסימום, ולכן ניתן להשמיט אותו.*
10. *הוצאת b החוצה מתוך הסכימה – b אינו תלוי במשתנה הסכימה. בנוסף, השמטת המינוס תוך חיפוש של הארגומנט שיביא את הביטוי למינימום ולא למקסימום – מעבר זה נכון מכיוון שהביטוי (b גדול ממש מ-0 וערך מוחלט תמיד גדול שווה 0).*
11. *הוא סקלר שאינו תלוי בארגומנט אותו אנו מחפשים שמביא את הביטוי למינימום, ולכן ניתן להשמיט אותו.*
12. *על פי חוקי ערך מוחלט.*
13. *על פי הגדרת התפלגות לפלס: .*
14. *הוא סקלר שאינו תלוי בארגומנט אותו אנו מחפשים שמביא את הביטוי למינימום, ולכן ניתן להוסיף אותו.*

**שאלה 3**

התשובה היא A.

B ו C לא, מכיוון שכל איטרציה מחלקת את המרחב בגרף לשניים ואין אפשרות לחלק את המרחב לשניים (בציר x ובציר y) בשאר הגרפים בלי שנקבל סתירה בנתונים כי לכל חלוקה שנעשה נקבל באותו צד פלוסים ומינוסים נכונים/שגויים וזה בלתי אפשרי.

D אפשר לחלק למשל ע"י כך שלכל *צריך להיות* ***(+)*** *ולכל צריך להיות* ***(-)****. אבל אנחנו מניחים תמיד שמשתמשים ב-ERM בשביל לאמן את ה-weak classifier. זאת אומרת שהוא חייב להיות decision stump שמחזיר את השגיאה הכי נמוכה על ההתפלגות הנתונה. וב D יש 3 טעיות וב A יש 2 טעיות (באמצעות הסיווג החלש שנציע בהמשך) אז האלגוריתם מעדיף את האפשרות עם הטעות הכי קטנה שזאת A.*

נציע *weak classifier* באופן הבא:

לכל *צריך להיות* ***(+)*** *ולכל צריך להיות* ***(-)****, בהנחה שכל משבצת בגרף היא יחידה אחת.*

*ואכן אפשר לראות שהאלגוריתם בגרף A נותן משקל יתר ל-* ***(+)*** *מכיוון שהם בצד הלא נכון של הגרף בשונה מה-* ***(-)****, שהם לא מקבלים משקל יתר מכיוון שהם בצד הנכון של החלוקה בגרף, ואכן אפשר לראות שיש רק 2 טעיות במקום 3 (בשונה מ D).*