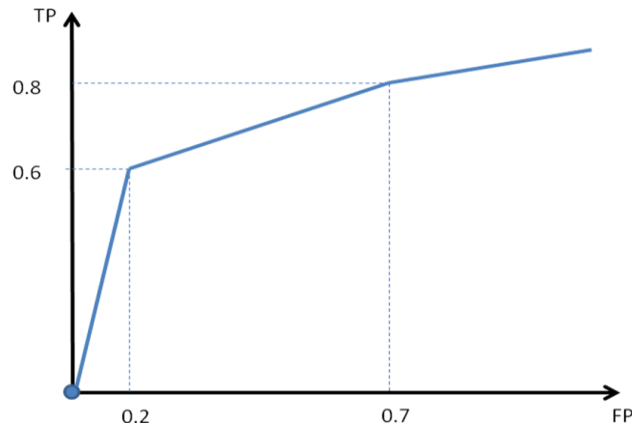


שאלה 1:

נניח שנתון לנו אלגוריתם למידה לגילוי "טקסטים העוסקים בנושאי ספורט", כלומר בהינתן טקסט - המחשב צריך להכריע האם טקסט זה עוסק בספורט או לא. בחנו את האלגוריתם על קבוצת מבחן במספר נקודות עבודה שונות וקיבלנו את הגרף הבא: TP כנגד FP:



אלו מהבאים הינו המתאים ביותר:

1. בחנו על 1000 טקסטים שמתוכם 500 עוסקים בנושא ספורט, בנקודת עבודה מסוימת קיבלנו זיהוי מדויק של 300 טקסטים כעוסקים בנושא ספורט עבור קבוצת ה"חיוביים", ובאותה נקודת עבודה 100 טקסטים מתוך קבוצת ה"שליליים" זהו כעוסקים בנושא ספורט.
2. בחנו על 1000 טקסטים שמתוכם 500 עוסקים בנושא ספורט, בנקודת עבודה מסוימת קיבלנו זיהוי מדויק של 100 טקסטים כעוסקים בנושא ספורט עבור קבוצת ה"חיוביים", ובאותה נקודת עבודה 300 טקסטים מתוך קבוצת ה"שליליים" זהו כעוסקים בנושא ספורט.
3. בחנו על 1000 טקסטים שמתוכם 500 עוסקים בנושא ספורט, בנקודת עבודה מסוימת קיבלנו זיהוי מדויק של 300 טקסטים כעוסקים בנושא ספורט עבור קבוצת ה"חיוביים" ובנקודת עבודה נוספת קיבלנו זיהוי מדויק של 350 טקסטים כעוסקים בנושא ספורט עבור קבוצת ה"חיוביים".
4. בחנו על 1000 טקסטים שמתוכם 500 עוסקים בנושא ספורט, בנקודת עבודה מסוימת קיבלנו זיהוי מדויק של 150 טקסטים כעוסקים בנושא ספורט עבור קבוצת ה"חיוביים" ובאותה נקודת עבודה 400 טקסטים מתוך קבוצת ה"שליליים" זהו כעוסקים בנושא ספורט.
5. בחנו על 1000 טקסטים שמתוכם 500 עוסקים בנושא ספורט, בנקודת עבודה מסוימת קיבלנו זיהוי מדויק של 300 טקסטים כעוסקים בנושא ספורט עבור קבוצת ה"חיוביים" ובאותה נקודת עבודה 350 טקסטים מתוך קבוצת ה"שליליים" זהו כעוסקים בנושא ספורט.

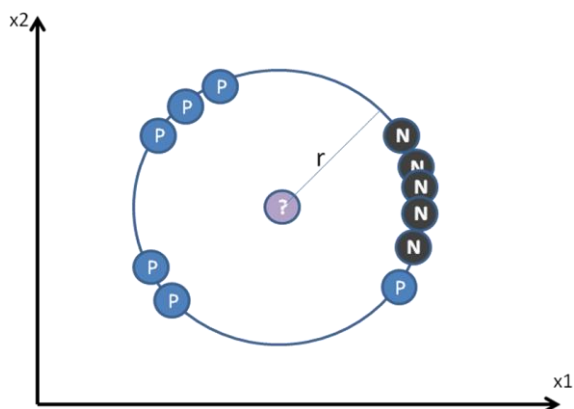
שאלה 2:

Cross-Validation

1. שיטה להערכת ביצועי אלגוריתם למידה
2. מונע בעיה של underflow עבור אלגוריתם Naïve Bayes
3. מציע לייצר עץ החלטה עם פיצולים רבים ככל שניתן.
4. אלגוריתם למידה דיסקרימינטיבי
5. אף תשובה אינה נכונה

שאלה 3:

נתונה קבוצת האימון הבאה:



תצפיות המסודרות ברדיוס r מסביב לנקודה אותה אנו רוצים לסווג. עפ"י אלגוריתם KNN ומטריקה L2, הנקודה שבמרכז תסווג:

1. P (אמנם כל הנקודות במרחק שווה אבל רוב העולם הוא P)
2. N (צפיפות הנקודות ב-N גבוהה יותר)
3. תלוי בשיטת משקול המרחקים
4. תלוי ב-K שנבחר
5. כל התשובות נכונות

שאלה 4:

נגדיר $E\text{-train}(h)$ כטעות של אלגוריתם למידה h על קבוצת האימון.

נגדיר $E\text{-test}(h)$ כטעות של אלגוריתם למידה h על קבוצת המבחן.

נאמר שלא אלגוריתם למידה h יש בעיה של overfitting אם קיים לנו אלגוריתם למידה h' ומתקיים:

1. $E\text{-train}(h) < E\text{-train}(h')$ אבל $E\text{-test}(h) < E\text{-test}(h')$
2. $E\text{-train}(h) < E\text{-train}(h')$ אבל $E\text{-test}(h) > E\text{-test}(h')$
3. $E\text{-train}(h) > E\text{-train}(h')$ אבל $E\text{-test}(h) < E\text{-test}(h')$
4. $E\text{-train}(h) > E\text{-train}(h')$ אבל $E\text{-test}(h) > E\text{-test}(h')$
5. כל התשובות נכונות

שאלה 5:

נתונה טבלת האימון הבאה:

| X1 | X2 | X3 | X4 | class |
|----|----|----|----|-------|
| 1 | 1 | 1 | 1 | P |
| 1 | 1 | 0 | 0 | P |
| 0 | 0 | 1 | 0 | N |
| 0 | 0 | 0 | 0 | N |

חשב את $IG(class|x2)$

1. 1
2. 0
3. 0.25
4. 0.75
5. לא ניתן לחשב

שאלה 6:

אלגוריתם SVM לינארי לעומת אלגוריתם פרספטרון:

1. אלגוריתם SVM בעל הכללה טובה יותר, קל לממש יותר פרספטרון
2. אלגוריתם SVM בעל הכללה טובה פחות, קל לממש יותר SVM
3. אלגוריתם SVM בעל הכללה טובה יותר, קל לממש יותר SVM
4. אלגוריתם SVM בעל הכללה טובה פחות, קל לממש יותר פרספטרון
5. רמת ההכללה בשני האלגוריתמים זהה, קל יותר לממש פרספטרון

שאלה 7:

נתון לנו ה-hyperplane לסיווג ווקטור בעל המאפיינים $f1, f2, f3$:

$$3 \times f1 + 2 \times f2 + 1 \times f3 + 0.5 \times f4 = T$$

כלומר, תוצאה גדולה/שווה ל-T ההכרעה היא "חיובי" ותוצאה קטנה מ-T הכרעה "שלילי".
יוסי הציע להגדיל את הערך של T. באופן עקרוני, פעולה זו גורמת ל-

1. הקטנת הערך של ה-TP והקטנת הערך של ה-FP
2. הגדלת הערך של ה-TP והגדלת הערך של ה-FP
3. הגדלת הערך של ה-TP והקטנת הערך של ה-FP
4. הקטנת הערך של ה-TP והגדלת הערך של ה-FP
5. אין משמעות לפעולה זו.

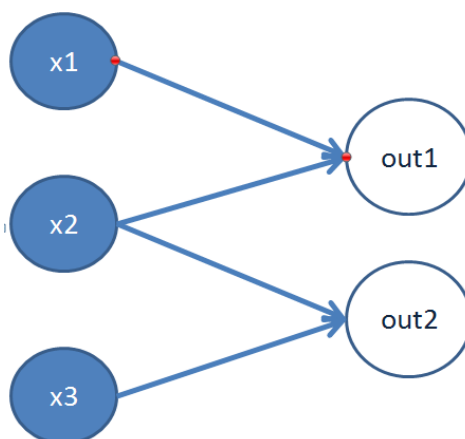
שאלה 8:

אלגוריתם Lazy

1. אינו מחשב מודל למידה בשלב ה"אימון", אלגוריתם KNN הינו אלגוריתם Lazy, חסרון גדול הינו הזמן לסיווג נקודה חדשה
2. אינו מחשב מודל למידה בשלב ה"אימון", אלגוריתם פרספטרון הינו אלגוריתם Lazy, חסרון גדול הינו הזמן לסיווג נקודה חדשה
3. מחשב מודל למידה כבר בשלב ה"אימון", אלגוריתם פרספטרון הינו אלגוריתם Lazy, חסרון גדול הינו זמן בניית המודל
4. מחשב מודל למידה כבר בשלב ה"אימון", אלגוריתם KNN הינו אלגוריתם Lazy, חסרון גדול הינו זמן בניית המודל
5. מחשב מודל למידה גם בשלב ה"אימון" וגם בזמן סיווג נקודה חדשה, אלגוריתם SVM הינו אלגוריתם Lazy, חסרון גדול הינו הזמן לסיווג נקודה חדשה.

שאלה 9:

נתונה לנו רשת הנוירונים הבאה:



המשקל של כל הקשתות הוא 1, ופונקציית האקטיבציה בשכבת היציאה הינה SGN עם ערך סף של 1 עבור כל אחת מהיציאות). מה יהיה הפלט של הרשת עבור הווקטור ($x_1=1, x_2=1, x_3=-0.5$)

1. Out1 = "POS", out2 = "NEG"
2. Out1 = "POS", out2 = "POS"
3. Out1 = "NEG", out2 = "POS"
4. Out1 = "NEG", out2 = "NEG"
5. לא ניתן לחשב

שאלה 10:

נתונה טבלת האימון הבאה:

| X1 | X2 | X3 | X4 | class |
|----|----|----|----|-------|
| 1 | 2 | 1 | 2 | 1 |
| 1 | 2 | 2 | 1 | 1 |
| 1 | 1 | 2 | 2 | -1 |
| 1 | 1 | 1 | 1 | -1 |

נדרש לסווג עם אלגוריתם Naïve-Bayes. המשתנים x_1, x_2, x_3, x_4 הינם בלתי תלויים זה בזה. רוצים לסווג את הווקטור $v = (x_1=1, x_2=2, x_3=2, x_4=2)$ עפ"י האלגוריתם, מהו $P(\text{class}=1 | v)$?

- 1. 0.6
- 2. 0.7
- 3. 0.25
- 4. 0.75
- 5. לא ניתן לחשב

שאלה 11:

הסדר הנכון בבניית מערכת מבוססת למידה ממוחשבת מונחית (Supervised Machine Learning) הוא (יש לבחור את ההיגד הנכון ביותר):

1. איסוף תצפיות, ייצוג הנתונים כוקטורים מתמטיים, תהליכי קדם-עיבוד (כגון נירמול, החלקה וכו'), בחירת אלגוריתם למידה, הערכת ביצועים, במידה שהביצועים לא מספקים יש לבחון כל אחד מהשלבים הקודמים האם יש צורך לשפרו
2. איסוף תצפיות, ייצוג הנתונים כוקטורים מתמטיים, תהליכי קדם-עיבוד (כגון נירמול, החלקה וכו'), בחירת מכונת למידה, הערכת ביצועים, במידה שהביצועים לא מספקים יש לבחור אלגוריתם למידה שונה
3. איסוף תצפיות, ייצוג הנתונים כוקטורים מתמטיים, תהליכי קדם-עיבוד (כגון נירמול, החלקה וכו'), בחירת מכונת למידה, הערכת ביצועים, במידה שהביצועים לא מספקים יש לשנות את תהליך הקדם-עיבוד
4. איסוף תצפיות, ייצוג הנתונים כוקטורים מתמטיים, תהליכי קדם-עיבוד (כגון נירמול, החלקה וכו'), בחירת מכונת למידה, הערכת ביצועים, במידה שהביצועים לא מספקים יש להוסיף תצפיות נוספות.
5. איסוף תצפיות, ייצוג הנתונים כוקטורים מתמטיים, תהליכי קדם-עיבוד (כגון נירמול, החלקה וכו'), בחירת מכונת למידה, הערכת ביצועים, במידה שהביצועים לא מספקים יש שנות את תהליך הקדם עיבוד.