מערכות לומדות תרגיל 2-א

שאלה 13:

הטבלה confusion matrix הנתונה היא בגודל 9x9, כך שכל i, כאשר מייצג ספרה המוצגת בתמונה. כלומר המחלקה אליה התמונה באמת שייכת על פי הערך של הספרה המצויה בתמונה. כל שורה היא ביחס לציר "true label". כל עמודה j בשורה i מייצגת את מספר הסיווגים השונים בערך j לתמונה שבפועל עם ערך i. בכל תא (i,j) כאשר i=j, מצוי ערך התמונות שהניבוי צדק ובשאר התאים הניבוי שגה.

שאלה 14:

כל עמודה היא ביחס לציר ״Predicted Label״, קרי כל עמודה j כאשר מייצגת את הערך שהמודל בפועל ניבה לתמונה מהסט. בכל שורה i בעמודה j ניתן לראות כמה מהניבויים עבור ערך j היו לתמונה שהכילה ערך i. בהתאמה לשאלה 13, כמות התמונות מהסט בעלות סיווג נכון נמצא בתא בו מס׳ השורה שקול למס׳ העמודה.

שאלה 15:

נסכום את כל הערכים המצויים בשורה מספר 3, נקבל 91 שזה הוא המכנה tp+fn, בתא (3,3) מצוי ערך 79=tp.

נסכום את כל הערכים במצויים בעמודה 9, נקבל 97 שזה הערך של המכנה tp+fp, בתא (9,9) מצוי ערך 90=tp.

שאלה 16:

עבור מעון ראש הממשלה אשר דורש ביטחון מקסימלי כלומר בדיקה מקיפה של ביטחון עם דיוק מוחלט, נעדיף להשתמש ב- מקסימלי. שכן נותן את אחוז הדיוק הכי גבוה בזיהוי הקלסיפיקציה.

שאלה 17:

מחלקת שיווק של חברה רוצה להתקשר למנויים שלה, על אף שהיא רוצה להתקשר לכולם היא לא יכולה לכן היא תרצה לדעת מי פוטנציאל לרכישה ולהתמקד בקבוצה זו. כלומר במקרה זה אנו נעדיף גבוה ולוותר על דיוק.

שאלה 18:

נזכיר את הטבלה של מסווג על פי הכלל ״תמיד בריא״:

|  |  |  |
| --- | --- | --- |
| clf - negative | clf - positive |  |
| 1 | 0 | true – positive p=1 |
| 999 | 0 | true – negative p=999 |

נזכיר את החישוב:

precision = tp / (tp + fp)

recall = tp / (tp + fn)

עבור positive קרי חולה:

בהתאמה עבור negative בריא:

התאמת הערכים היא ביחס למחלקה אותה אנו מחשבים – חולה/בריא ולכן הערכים שונים. חישבנו את הערכים בהתאם לתיאור החישוב בשאלה 15.