תרגיל מס' 1

מצורפים שני קבצים עם 50000 דוגמאות של dataset ידוע שנקרא CIFAR10. הקובץ הראשון מכיל את הפיטצ'רים (16 פיטצ'רים לכל תמונה, שאני הוצאתי במיוחד בשבילכם) הקובץ השני מכיל את מספר המחלקה (מספר בין 0 ל-9).

מוזמנים להסתכל: https://www.cs.toronto.edu/~kriz/cifar.html

המטרה המרכזית של התרגיל היא לכתוב קוד של אחד נגד כולם.

אתם מוזמנים להשתמש ב LogisticRegression מוכן, בספריית sklearn . שימו לב שיש למחלקה שתי אפשרויות אימון אחת בשיטת softmax ואחת אחד נגד כולם.

1. תשתמשו בשתי השיטות והשוו את אחוזי ההצלחה ואת זמני הריצה.
2. לכל שיטה, חשבו את ערך פונקציית המחיר
3. לכל שיטה, חשבו את F1-mean
4. עבור שיטת האחד נגד כולם, הביטו במטריצת הבלבול ובדקו האם יש צמדי מחלקות שיותר קשה להבדיל ביניהם
5. עבור זוג אחד כזה, תיצרו מודל נוסף, שמבדיל רק בין 2 המחלקות. השתמשו בו למקרה בו קבלתם במודל הראשון את אחת משתי המחלקות. האם התוצאות השתפרו?

תחלקו ל 70% ללמידה ו-30% לבדיקה.

יש להגיש רק קובץ ipynb (מותר לכם לכל תא לסמן שהוא הערות ולא קוד כך תוכלו לרשום מסקנות)

בקוד צריכה להיות פונקציה testmymodel שמקבלת מודל, קובץ numpy של פיטצ'רים וקובץ numpy של מחלקות ומחזיר את אחוזי ההצלחה של המודל! אתם תיבדקו על דאטה שלא קיבלתם כלל (בתוך הפונקציה מותר לכם גם להכניס חישובים נוספים כרצונכם – לא חייבים אפשר רק predict ו-score)

תוכלו להיעזר בפקודה:

|  |
| --- |
|  |
|  | X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, Y, test\_size=0.3, random\_state=42) |