

# סיווג ספרות באמצעות sklearn ו-GDA

## להגשה במסגרת תרגיל בית

בבעיה זאת תאמנו 3 מסווגים ממשפחת GDA לסיווג תמונות של ספרות (כתב יד), תשתמשו בהם לסיווג הדוגמאות, ותיצרו סיכום מפורט של ביצועיהם. **הערה:** בכל המקומות בהם מוצג למטה קוד הוא ממורף כהצעה, והינכם רשאים לעשות בו שימוש או לחלופין לכתוב קוד אחר ע"פ שיקול דעתכם.

### חלק א – הכנות

1. צרו מחברת בשם GDAForHandwrittenDigitsClassification.ipynb

2. טעינת המידע

2.1. טענו את המידע מהקובץ features.csv ושמרו אותו כמטריצה בשם X

2.2. טענו את המידע מהקובץ labels.csv ושמרו אותו כוקטור בשם y

3. הציגו בגרף את 50 השורות הראשונות במטריצה X, והסבירו בקצרה מה ניתן לראות בגרף שיצרתם

```
plt.imshow(X[:50], cmap='gray');
```

4. חלקו את המידע לסדרת אימון וסדרת מבחן ביחס של 1:4

```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state = 0)
```

### חלק ב – אימון מסווגים ומדידת ביצועיהם

5. השתמשו ב sklearn בכדי לאמן מסוג מסוג QDA

```
from sklearn.discriminant_analysis import QuadraticDiscriminantAnalysis
```

6. מדדו ודווחו עבור המסווג שאימנתם את הגדלים הבאים, כל אחד מהם עם סדרת האימון וסדרת המבחן בנפרד

6.1. דיוק הסיווג

6.2. מטריצת הערבול

```
from sklearn.metrics import confusion_matrix
```

6.3. עבור כל תיוג בנפרד, את ה precision, recall, f1-score

```
from sklearn.metrics import classification_report
```

7. חזרו על 5+6 עבור מסוג מסוג LDA

```
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
```

8. חזרו על 5+6 עבור מסוג מסוג GNB

```
from sklearn.naive_bayes import GaussianNB
```