

## פרק 3 : סיווג.

### חלק תיאורטי:

#### 1. מבוא:

קרא מתוך ISLR עמודים 127-130.

#### 2. Logistic Regression:

קרא מתוך ISLR עמודים 130-138.

#### 3. Naïve Bayes:

קרא את המאמר המצורף: Naïve Bayes Explained .

#### 4. LDA:

קרא מתוך ISLR עמודים 138-150.

#### 5. KNN:

לממש KNN וקטורי בעצמך ולהשוות ביצועים למודל הממומש בSKlearn.

לצורך התרגיל הנ"ל לקחנו את הקוד מהתרגיל הראשון בקורס 231CS של סטנפורד. בו יש מחברת בשם KNN אשר מכילה פונקציה TODO בה יש להשלים את פעולת הKNN בצורה ווקטורית.

ניתן לקרוא הרחבה של KNN במבוא בקישור הבא(להיזהר לא לקרוא את התשובות ;) :

<https://kevinzakka.github.io/2016/07/13/k-nearest-neighbor/>

#### 6. סיכום:

קרא מתוך ISLR עמודים 151-154.

### חלק מעשי:

בתרגיל זה נשתמש ב-iris dataset המצורף לתיקייה זו.

הנתונים בקובץ הנ"ל מתחלקים לשלושה סוגי אירוסים: Setosa, Versicolour, Virginica.

לכל רשומה ארבעה מאפיינים:

- אורך עלי הגביע (Sepal Length).
- רוחב עלי הגביע (Sepal Width).

- אורך עלי הכותר (Petal Length).
- רוחב עלי הכותר (Petal Width).

עליך לבנות מסווג בעל דיוק מירבי לסיווג אירוסים לשלוש הקטגוריות הנ"ל.

בפתרון עליך לכלול השלבים הבאים:

- אקספלורציה של המידע.
- בחינת מודלים.
- השוואה בין המודלים שנבחנו.
- מסקנות.