

## כריית נתונים - מעבדה 1 - עיבוד מקדים

במעבדה זו יש לממש בפייתון על גבי ג'ופיטר, פונקציות שונות העובדות על סדרת נתונים, אין להשתמש בפונקציות או ספריות מוכנות - יש לממש בעזרת פייתון והספריות הסטנדרטיות בלבד. (ניתן להעזר במקורות המצורפים, מצופה גם לבצע חיפוש ובדיקה עם מקורות מידע נוספים מהרשת)

### 1. נירמולים

- יש לסכם בקצרה מהו נרמול נתונים, מהי סטנדרטיזציה של נתונים וההבדל ביניהם
- יש לסכם בקצרה על כל סוג נרמול/סטנדרטיזציה (סעיפים a-c)
- יש לממש פונקציה המקבלת סדרת נתונים ופרמטרים רלוונטיים ומחזירה אותה לאחר נרמול בהתאם לפונקציה המתבקשת (עבור כל סעיף פונק נפרדת)
- יש לחקור ולמצוא ספריות פייתון שיודעות לבצע נרמול/סטנדרטיזציה ובצע בעזרתן את אותן פעולות

- a. Min-max
- b. Z-score
- c. Decimal Scaling

מקורות לקריאה :

[https://scikit-learn.org/stable/auto\\_examples/preprocessing/plot\\_all\\_scaling.html](https://scikit-learn.org/stable/auto_examples/preprocessing/plot_all_scaling.html)  
<https://medium.com/@dataakkadian/standardization-vs-normalization-da7a3a308c64>  
<https://www.geeksforgeeks.org/data-normalization-in-data-mining/>

### 2. דיסקרטיזציה לא מונחית של ערכים רציפים - Discretization using binning

- הסבירו מהי דיסקרטיזציה
- מהי דיסקרטיזציה של עומק שווה Equal-frequency discretization ממשו פונקציה המקבלת סדרת מספרים ופרמטרים רלוונטיים ומחזירה סדרה של ערכים בדידים כנדרש
- מהי דיסקרטיזציה של רוחב שווה Equal-width discretization ממשו פונקציה המקבלת סדרת מספרים ופרמטרים רלוונטיים ומחזירה סדרה של ערכים בדידים כנדרש
- יש לחקור ולמצוא ספריות פייתון שיודעות לבצע דיסקרטיזציה בעזרת דליים ולבצע בעזרתן את אותן פעולות
- 
- [https://www.saedsayad.com/unsupervised\\_binning.htm](https://www.saedsayad.com/unsupervised_binning.htm)

### 3. החלקה

- a. הסבירו מהי החלקת נתונים (smoothing)
- b. הסבירו מהו ממוצע נע ואיך הוא עוזר בהחלקת נתונים
- c. פרטו על השיטות וממשו פונקציות המבצעות את ההחלקה
  - i. Simple Moving Average
  - ii. Weighted Moving Average
  - iii. Exponential Moving Average
- d. הסבירו על Binning Methods for Data Smoothing
  - i. Smoothing by bin means:
  - ii. Smoothing by bin boundaries

יש לחקור ולמצוא ספריות פייתון שיודעות לבצע החלקה ולבצע בעזרתן את אותן פעולות

<https://www3.nd.edu/~rjohns15/cse40647.sp14/www/content/lectures/06%20-%20Data%20Quality.pdf>

[https://t4tutorials.com/binning-methods-for-data-smoothing-in-data-mining/#Binning\\_Methods\\_for\\_Data\\_Smoothing](https://t4tutorials.com/binning-methods-for-data-smoothing-in-data-mining/#Binning_Methods_for_Data_Smoothing)

<https://machinelearningmastery.com/moving-average-smoothing-for-time-series-forecasting-python/>

<https://www.sciencedirect.com/topics/economics-econometrics-and-finance/smoothing-technique>

<https://www.itl.nist.gov/div898/handbook/pmc/section4/pmc42.htm>