



תרגיל 7 – Linear Regression

הוראות הגשה:

1. בתרגיל הבא יש לענות על השאלות באמצעות שימוש בקוד פייתון ושימוש ב-Scikit-Learn.
2. יש להגיש את העבודה בזוגות בלבד.
3. התרגיל יוגש כמחברת colab, כאשר לתיבת הגשה יש להגיש קובץ המכיל לינק למחברת עם הרשאות קריאה למייל yaelhoc@post.bgu.ac.il + tamarin@post.bgu.ac.il
4. שם הקובץ יהיה מספרי הזהות של המגישים בצורה הבאה: זהות1_זהות2
במחברת הפתרון, יש לציין את מספר השאלה עליה עניתם עבור כל חלק בפתרון

ענו על השאלות הבאות באמצעות הנתונים על מחירי יהלומים:

1. חשבו קורלציה של Pearson בין עמודת carat לבין עמודת המחיר
2. חשבו מטריצת קורלציות בין כל העמודות הנומריות (עמודות עם ערכים מספריים) וציירו את המטריצה באמצעות חבילת seaborn.
3. עבור כל סוג cut, צרו מודל רגרסיה לינארית פשוטה עם משתנה אחד שחוזר באמצעות נתוני carat את מחיר היהלום.
4. עבור כל מודל רגרסיה שיצרתם בשאלה 3, הציגו את קו הרגרסיה שהתקבל בדיאגרמה אחת עם מקרא (legend) כאשר לכל קו יש צבע אחר לפי ערך cut.
5. חלקו את הנתונים לtest וtrain ובנו רגרסיה מרובת משתנים, שחוזר את מחיר היהלום באמצעות כל העמודות הנומריות. חשבו מה ערכי MAE ו RMSE של המודל שיצרתם.

ענו על השאלות הבאות באמצעות הנתונים של מכירת רכבים בארה"ב:

1. עבור כל הרכבים של ford, בנו מודל רגרסיה לינארית פשוטה, שמחשב את מחיר הרכב לפי mileage. ציירו את המודל של הרגרסיה שבניתם.
 2. עבור כל הרכבים של dodge, בנו מודל רגרסיה לינארית מרובת משתנים, שמחשבת את מחיר הרכב לפי mileage ולפי עמודת ה-Year.
- בנוס (5 נק'): באמצעות חבילת plotlyexpress ציירו את המודל שהתקבל בתלת מימד.