



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. שם הקובץ צריך להיות: ID1_ID2.ipynb, סוג הקובץ שאתם מורידים מהcolab הוא ipynb.ipynb.
4. במחברת שאתם מגישים תאי הפלט צריכים להראות את הפלט שהצתם. אין להגיש קוד בלבד ללא הפלט שלו.

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 עמודות carat ו depth 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 clarity, צרו מודל רגרסיה לינארית פשוטה עם משתנה אחד שחוזר באמצעות נתוני הcarat 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 הcut.
5. חלקו את הנתונים לtrain וtest 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 Toyota ו dodge (בנפרד), בנו מודל רגרסיה לינארית פשוטה, שמחשב את מחיר הרכב לפי mileage. ציירו את המודל של הרגרסיה שבניתם.
2. עבור כל הרכבים של Ford, בנו מודל רגרסיה לינארית מרובת משתנים, שמחשבת את מחיר הרכב לפי mileage ולפי עמודת ה-Year.
3. באמצעות חבילת plotlyexpress ציירו את המודל שהתקבל בדו מימד ובתלת מימד.

ענו על השאלות הבאות באמצעות הנתונים של המיליארדים לפי מגזין פורבס:

1. עבור כל industry שקיימים בהם יותר מ-3 מיליארדים חשבו קורלציה של NetWorth לבין עמודת הגיל
2. הציגו בצורה ויזואלית את התוצאות שקיבלתם בסעיף 1.