

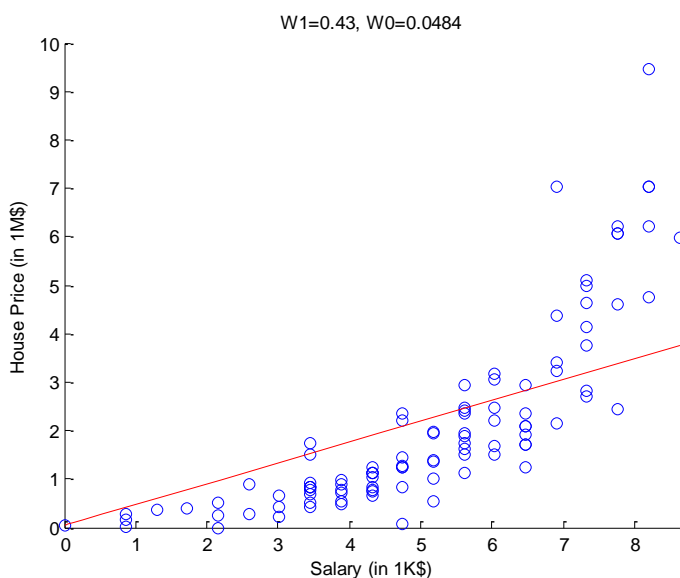
תרגיל מס' 2: רגרסיה לינארית, KPI

תאריך אחרון להגשה: 19/12/2019

רגרסיה לינארית

כתבו קוד מטלב אשר מממש רגרסיה לינארית.

השתמשו בקובץ הנתון LinearRegression.m כשלד. יש לעבור על הקובץ בעיון, ולוודא שאתם מבינים את קוד המטלב שבו. בקוד יש הפניה לשלוש פונקציות:
1. `plotWithPrediction` – פונקציה המציירת תרשים המתאר את הנתונים כוללת `scatter` של כל הנקודות הנתונות וכן קו המתאר את הפונקציה החזויה. דוגמא לתרשים שהפונקציה מוציאה:



2. `doBatchGD` – הפונקציה מבצעת איטרציה אחת של Gradient Descent ב `Batch`. עליכם להשלים את הקוד על פי ההערות הנתונות בקוד.
3. `predictTestData` – פונקציה המשתמשת במשקולות שהאלגוריתם מצא על מנת לחזות את ערכי הנתונים מה `Test Set`.

בכל אחת מהפונקציות הנ"ל עליכם להשלים את הקוד החסר (שימו לב להערות שמתחילות במילים: "Add Code"). אין צורך לשנות את הקוד הראשי שבקובץ `LinearRegression.m`

לאחר שהקוד שלכם רץ ועובד ענו על השאלות הבאות:

1. כמה איטרציות לקח לאלגוריתם שלכם להתכנס? פרטו כיצד הגעתם לתוצאה זאת?
2. לאחר סיום הרצת האלגוריתם, איזו שגיאה גדולה יותר – השגיאה עבור סט האימון או השגיאה עבור סט הבדיקה? האם התוצאה צפויה? מדוע?

3. קצב הלמידה: שנו את קבוע הלמידה לערכים הבאים: 0.1, 0.01, 0.0001. עבור כל אחד מהערכים הנ"ל הריצו את הקוד וענו על השאלות הבאות
- מה השתנה באלגוריתם? מדוע?
 - איזה מהקבועים הנ"ל (בתוספת הקבוע המקורי – 0.001) הוא הטוב ביותר עבור הנתונים?
4. הביטוי בנתונים, האם לדעתכם הפונקציה שמצאנו מסבירה את הנתונים בצורה טובה? אם לא, מדוע?

KPI

בקובץ KPI_data שמורים הערכים האמיתיים של אלף דוגמאות בוקטור `realValues`, לכל דוגמה יש ערך של 1 (positive example) או 0 (negative example). כמו כן שמורים בקובץ גם קבצי הפלט משלושה קליסיפיירים שונים, כך שכל איבר בוקטורי הפלט הוא החיזוי של הקליסיפיייר עבור הדוגמה המתאימה מ-`realValues`. החיזויים הם כולם ערכים בין 0 ל-1, יש להתייחס לחיזוי שערכו מעל 0.5 כחיזוי חיובי ולחיזוי שקטן או שווה ל-0.5 כחיזוי שלילי.

עבור כל קליסיפיייר יש לחשב `accuracy`, `recall`, `precision`, `F-Measure`, ולהסביר בעזרת האינדיקציות שחושבו מה היתרון של הקליסיפיייר (אם קיים).

ניתן להשתמש בקובץ KPI כשלב לפיתרון.

התשובה שלכם עבור כל קליסיפיייר אמורה להיראות בערך ככה:

ערכי ה-KPI שחושבו עבור הקליסיפיייר הראשון הם:

Accuracy=...

Recall=...

Precision=...

F-Measure=...

הייתי משתמש בקליסיפיייר זה במקרה שהייתי רוצה...

תזכורת:

- שימוש ב-`>`, `<`, `==` וכו' על וקטורים מחזיר וקטור בוליאני באותו אורך שבו יש 1 עבור כל איבר בוקטור המקורי שמקיים את התנאי ו-0 עבור כל איבר שלא מקיים את התנאי.
- שימוש ב-`&` מאפשר קבלת וקטור בוליאני שבו ערך 1 עבור איברים שמקיימים כמה תנאים. (למשל השורה `a>0.5 & b==1` תחזיר וקטור בוליאני שבו 1 עבור כל מקרה שבו האיבר המתאים ב-a היה גדול מ-0.5 וגם האיבר המתאים ב-b היה שווה ל-1, בהנחה ש-a ו-b הם וקטורים באותו אורך.)
- ניתן להשתמש בפונקציה `sum` על מנת לסכום את מספר האיברים בוקטור בוליאני השווים לאחד, כלומר `sum(a>0.5 & b==1)` ייתן את מספר האיברים שבורם הערך ב-a גדול מ-0.5 והערך ב-b שווה ל-1.

יש להגיש את קובצי הקוד עם השינויים שלכם, וכן קובץ `word` עם התשובות לשאלות כולל גרפים והסברים היכן שצריך.

בהצלחה!