

17



מס. נבחן: 8909581
7029210-1
חדר: 60-1.5
ת.ז. 316095660

המחלקה
למדעי המחשב

אוניברסיטת
אריאל
בשומרון

מבחן במסדי נתונים

עמוס עזריה ומירב שקרון

7029210-1,3,4,5

סמסטר ב' מועד א' ג' באב התשפ"א, 12.7.2021

הנחיות כלליות:

- משך הבחינה: 150 דקות.
- יש לענות בגוף השאלון! המחברת תשמש כטיוטא בלבד, מענה במחברת עלול לגרור ציון 0.
- אין להכניס שום חומר עזר.
- השימוש במחשבון אסור.
- בשאלות האמריקאיות רק תשובה אחת נכונה. שימו לב שיש לנמק את הבחירה
- בסיום הבחינה - נא למסור את השאלון ואת המחברת.

	1	2	3	4	5	6	Total
Max points	28	19	10	15	15	18	105
Grade	28	19	7	14	15	18	101

בהצלחה!



שאלה 1-SQL (28 נק')

נתונות הטבלאות הבאות המתארות נתונים מקופת חולים מסוימת הקשורים למחלת הקורונה:

patients(id, first_name, last_name, birth_date, gender, first_dose_date, second_dose_date)

[טבלת מטופלים: תז, שם פרטי, שם משפחה, תאריך לידה, מגדר, תאריך מנת חיסון ראשונה, תאריך מנת חיסון שניה]

isolated(patient_id, start_date, end_date, isolation_type_id)

[טבלת מבודדים: תז, תאריך כניסה לבידוד, תאריך סיום הבידוד, מזהה סוג בידוד]

isolation_types(id, desc)

[טבלת סוג הבידוד: מזהה סוג בידוד, תיאור סוג בידוד]

הערכים בטבלה זו הם:

id	desc
1	חדר ושירותים נפרדים. בבית גרים אנשים נוספים
2	חדר נפרד בבית. שירותים עם שאר יושבי הבית
3	לבד בבית
4	מלונית

confirmed_positive_cases(patient_id, start_date, is_symptomatic, end_date)

[טבלת חולים מאומתים: תז, תאריך תחילת מחלה, האם יש סימפטומים, תאריך סיום המחלה]

between

א. כתבו שאילתא שמחזירה את פרטי המטופלים העונים על כל התנאים הבאים:

- חלו לאחר שהתחסנו במנה ראשונה (ולא השנייה)
- ✓ כבר החלימו מהמחלה
- ✓ הם בקבוצת הגיל 20-33

עבור כל מטופל/ת יש להציג את הנתונים הבאים:

- שם פרטי
- שם משפחה
- תז
- מס' הבדיקות שעשה/תה אי פעם

(14 נק')

Select p.first-name, p.last-name, p.id
 From (patients as p join confirm-positive-cases as c) AS all
 on p.id = c.patient_id
 where p.birth-date > newdate(7, 12, 1989) And
 p.birth-date < newdate(7, 12, 2001) And
 c.end-date < newdate(7, 12, 2021) And
 c.start-date > p.first-dose-date And
 c.start-date < p.second-dose-date

(שמות א"ת וס' (טקסט) המיון (א) < י"ב קטן קטן)

הסבר - נבחר בין שתי החלוקות ונבחר ונבחר אלו שיש להם שתי מחסנים
 אלו הם חלק מהאוכלוסיה ואלו הם חלק מהאוכלוסיה.

ב. כיתבו שאילתא שתציג את מס' האנשים שחלו בתקופת הבידוד שלהם לפי סוג הבידוד. כלומר, לכל סוג בידוד ייספרו רק מטופלים שאומתו בזמן שהותם בבידוד.

בכל רשומה יש להציג:

- מזהה סוג הבידוד
- תיאור סוג הבידוד
- מס' המטופלים שחלו בזמן שהותם בבידוד זה

(14 נק')

Select Itype.id, Itype.desc, count(*)
 From (confirm-positive-cases as c Join isolated as I
 on c-patient-id == I-patient-id) Join isolated-types as Itype
 on c-isolation-type-id == Itype.id
 Where c.start-date > I.start-date And
 c.start-date < I.end-date
 Group by Itype.id

הסקר נחקר 3 מלאות - המובנים, סוג מרכז וחלל

להתייחס (המובנים וחלל) ונראה שהם אותו חלל בתקופת הבידוד
 ולא נראה מספרים אחרים Group by על כל המידות נדרש.

2. Normalization (19 נק')

נתונה הרלציה הבאה:

 $R(A, B, C)$

נתונות השורות (הרשומות) הבאות ברלציה:

 a_1, b_2, c_1 a_1, b_1, c_2 a_2, b_2, c_1

כמו כן, נתון שמהחוקיות של הרלציה, ומהשורות הנתונות נובעות גם השורות הבאות (ייתכן שנובעות גם שורות נוספות):

 a_1, b_1, c_1 a_1, b_2, c_2

א. מה ניתן לומר על NF של הרלציה? נמקו היטב. (10 נק')

נמצא שיש שתי שורות - NF - ראשית שכל חלק מהשורה הוא חלק מהשורה האחרת.
 NF - שני שורות שחלק מהשורה הוא חלק מהשורה האחרת.
 NF - שתי שורות שחלק מהשורה הוא חלק מהשורה האחרת.

→ 3.5 NF - בואו שיש שתי שורות שחלק מהשורה הוא חלק מהשורה האחרת.
 ג. שתי שורות שחלק מהשורה הוא חלק מהשורה האחרת.

4 NF = לא מתקיים. ניתן להראות שיש שתי שורות שחלק מהשורה הוא חלק מהשורה האחרת.
 הראשונה השנייה והשלישית.
 הראשונה השנייה והשלישית.
 הראשונה השנייה והשלישית.
 הראשונה השנייה והשלישית.

באופן כללי
 ראשית
 שני שורות שחלק מהשורה הוא חלק מהשורה האחרת.
 שתי שורות שחלק מהשורה הוא חלק מהשורה האחרת.
 שתי שורות שחלק מהשורה הוא חלק מהשורה האחרת.

ב. עבור כל אחת מהשורות להלן, ציינו האם ייתכן שהיא ברלציה (כיתבו "ייתכן"), לא ייתכן שהיא ברלציה (כיתבו "לא ייתכן"), או שהיא מוכרחת להיות ברלציה כיוון שהיא נובעת מכל הנתונים (כיתבו "נובעת"):

(9 נק')

a_2, b_2, c_2 ייתכן
 a_2, b_1, c_2 ייתכן
 a_2, b_1, c_1 ייתכן

הצורה:

אם הבנתם וזכור את השאלה אם הכוונה היא שיש שתי שורות שחלק מהשורה הוא חלק מהשורה האחרת.
 אם כן, אז לא סתם את השורות הראשונה והשנייה שחלק מהשורה הוא חלק מהשורה האחרת.
 אם כן, אז לא סתם את השורות הראשונה והשנייה שחלק מהשורה הוא חלק מהשורה האחרת.
 אם כן, אז לא סתם את השורות הראשונה והשנייה שחלק מהשורה הוא חלק מהשורה האחרת.

or
שם

//
[שם]

6

3. XPath – XML (10 נק')

נתון מסמך xml המתאר נתונים על חולי קורונה ומבודדים.
להלן דוגמא למבנה ותכולת הxml-

```
<corona>
  <isolations>
    <isolated>
      <id>335598740</id>
      <full_name>"Tom Segev"</full_name>
      <age>36</age>
      <gender>male</gender>
      <start_date>01/10/2021</start_date>
      <end_date>14/01/2021</end_date>
      <place>"חדר ושירותים נפרדים. בבית גרים אנשים נוספים"</place>
    </isolated>
    .....
  </isolations>
  <confirmed_positive_cases>
    <case>
      <id>775521547</id>
      <name>"Jacky Levi"</name>
      <age>56</age>
      <gender>male</gender>
      <start_date>10/10/2021</start_date>
      <is_symptomatics>"True"</is_symptomatics>
    </case>
    .....
  </confirmed_positive_cases>
</corona>
```

כתבו שאילתא אחת ב-XPath שתחזיר את שמות החולים והמבודדים שגילם גדול מ-40 (מטופלים שמופיעים גם ברשימת החולים וגם ברשימת המבודדים יופיעו פעמיים בתוצאה)

//corona/isolations/isolated/full_name [age > 40]

//corona/confirmed-positive-cases/case [age > 40] Fullname

הערה: // - אומר לך לא כל המידע בלבד, אלא כל המידע שיש לו את השם הזה.
הערה: [age > 40] - אומר לך לך את כל המידע שיש לו את השם הזה ויש לו גיל מעל 40.

אם אתם רוצים לקבל את כל המידע, אז אתם צריכים לכתוב: //corona/isolations/isolated/full_name

4. MongoDB (15 נק')

נתון collection של נתונים ב-MongoDB המאחסן נתוני מתחסנים לפי ערים בארץ.

שם ה-collection: Vaccinated.

להלן דוגמא למסמך ב-collection:

```
{
  _id: ObjectId("50a8240b927d5d8b5891743c"),
  id: 0335598750
  city: "Ariel",
  first_name: "Yael",
  last_name: "Cohen",
  gender: "female",
  age: "39"
  first_dose_date: new Date("May 04, 2021"),
  second_dose_date: new Date("May 25, 2021"),
  is_recovered: "True"
}
```

כתבו שאילתא ב-MongoDB שתחזיר לכל עיר את מס' המחוסנים (חיסון ראשון) שהם גם מחלימים וגם מעל גיל

30.

mapReduce

function()

if (is_recovered == 'True') { if (age > 30) { if (first_dose_date < new Date('July 1, 2021'))

emit(city, id);

function(a,b)

return {a, b.size()};

}

(map reduce) זהו סוג של פונקציה

המבצע mapReduce מחלק את הנתונים לקבוצות קטנות יותר ומבצע על כל אחת מהן פונקציית map, ולאחר מכן מבצע פונקציית reduce על תוצאות המפה.

5. Java Streams (15 נק')

כיתבו באמצעות שימוש ב-Java Streams בלבד פקודה המדפיסה את כל המספרים הראשוניים מ-2 ועד X נתון. אין לבנות פונקציית עזר, אין להשתמש בפונקציה קיימת לבדיקת ראשוניות (כגון: isProbablePrime) אין צורך להתחשב ביעילות. (רמז: העזרו ב- `IntStream.rangeClosed(a,b)` המחזירה stream המכיל את כל המספרים מ-a ועד b).

```
int X;

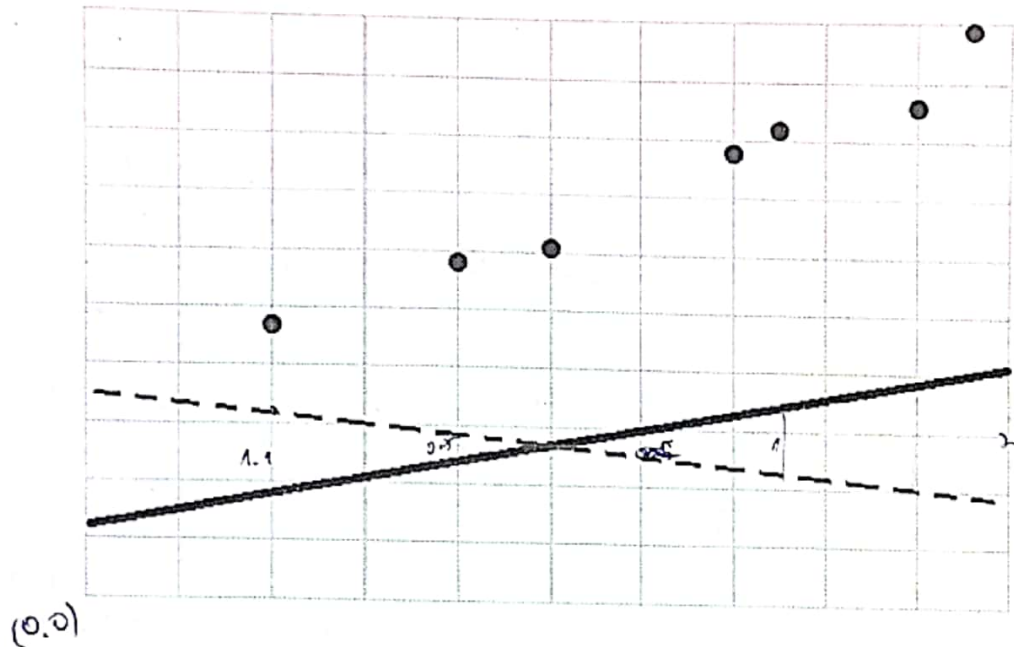
IntStream.rangeClosed(2, X).filter(a -> a % (IntStream.rangeClosed(2, a-1))
    == 0).forEach(System.out.println);
```

15

המסר: נ'צני stream על כל n כל 2 עד X נתון.
 נעשה עליו filter וניסיון לנו את n מה ~~המספרים~~ הנתונה
 והמספרים עליו (x) שונה מ-0 ולכן ציינו לקבוע ~~אם~~ n
 לה נאסוף n שצין לקבוע n המספרים n - 2 עד
 המספר הנספר מתוך 1 ולכן ישל כל המספרים הנאספים
 נעשה את ה"קולטה" ונו בליון את stream.
 בסוף ציין מהפס אל ונעשה `forEach` על פקודת הדפסה.

6. Linear regression (18 נק')

נתון הגרף הבא המתאר דאטה (באמצעות הנקודות) עם פיצ'ר בודד (בציר ה-x) וה target label (בציר ה-y), וכן שני מודלים לינאריים: הקו הרציף והקו המרוסק. הניחו שהקצה השמאלי למטה הוא ראשית הצירים (0,0).



ענו על השאלות הבאות. נמקו את תשובותיכם!

א. איזה מהמשפטים הבאים נכון בנוגע ל Loss (פונקציית הטעות) של שני המודלים: (6 נק')

- a. Loss של המודל של הקו המרוסק נמוך יותר מזו של הקו הרציף.
- b. Loss של המודל של הקו הרציף נמוך יותר מזו של הקו המרוסק.
- ג. לא ניתן לדעת איזה Loss נמוך יותר כיוון שאין מספרים על הצירים.
- d. Loss של שני המודלים זהה, שכן כל הנקודות נמצאות מעל הקווים.

פונקציית ה Loss מחשבת את סכום הריבועים של ההפרשים בין הערכים המצויים לאלה שהמודל חז'ה. אין צורך להעריך את הפונקציה שירותי ניתן/אופן כבידור מהפנטים בין שני הערכים המצויים לאלה שהמודל חז'ה. אין צורך להעריך את הפונקציה שירותי ניתן/אופן כבידור מהפנטים בין שני הערכים המצויים לאלה שהמודל חז'ה.

18

ב. נסתכל על המודל של הקו המרוסק: מה ניתן לומר על w של המודל של הקו המרוסק, ומה יקרה לה באיטרציה הבאה של Gradient descent? (6 נק')

- (a) w הנוכחית שלילית והיא תגדל (כלומר תהיה פחות שלילית או תהפוך חיובית).
 b. w הנוכחית שלילית והיא תקטן (כלומר תהיה יותר שלילית).
 c. w הנוכחית חיובית והיא תגדל.
 d. w הנוכחית חיובית והיא תקטן (כלומר תהיה קטנה יותר או תהפוך שלילית).
 e. לא ניתן לדעת האם w תקטן או תגדל לפי הנתונים בשאלה, הדבר תלוי במה שקרה באיטרציה הקודמת.

לפי בנקודות הקו המרוסק הוא נמצא ב- $w(x) = w_0 + w_1x$ ניתן לראות
 ומבנקודות יורדת וקו הוא שלילי.
 במבט על הנקודות המצויים נראה כי הקו הוא שלילי ונמצא
 וקו כולו, שקודקודו הוא w_0 וציר w_1 הוא חיובי, ולכן הנקודה
 המצויה היא חיובית.

ג. נסתכל על המודל של הקו הרציף: מה ניתן לומר על הנגזרת של Loss בנקודה w, b התואמים את אלו של המודל של הקו הרציף. זכרו שב- Gradient descent אנו תמיד מתקנים בצעד בכיוון הפוך לנגזרת. (6 נק')

- a. לא ניתן לומר דבר כיוון שאין מספרים על הצירים.
 b. הנגזרת החלקית לפי w בנקודה w, b תהיה חיובית, ואילו הנגזרת החלקית לפי b בנקודה w, b תהיה שלילית.
 c. הנגזרת החלקית לפי w בנקודה w, b תהיה שלילית, ואילו הנגזרת החלקית לפי b בנקודה w, b תהיה חיובית.
 d. הנגזרת החלקית לפי w בנקודה w, b תהיה חיובית, וכן הנגזרת החלקית לפי b בנקודה w, b תהיה חיובית.
 (e) הנגזרת החלקית לפי w בנקודה w, b תהיה שלילית, וכן הנגזרת החלקית לפי b בנקודה w, b תהיה שלילית.

הנגזרת של הפונקציה w היא חיובית, ולכן נקט צעד שלילי, מה שמשנה את w ל-
 $w = w - a \cdot \frac{\partial L}{\partial w}$
 (כיוון של w הוא חיובי, $\frac{\partial L}{\partial w}$ הוא שלילי, ולכן w יגדל).
 הנגזרת של הפונקציה b היא שלילית, ולכן נקט צעד חיובי, מה שמשנה את b ל-
 $b = b - a \cdot \frac{\partial L}{\partial b}$
 (כיוון של b הוא שלילי, $\frac{\partial L}{\partial b}$ הוא חיובי, ולכן b יגדל).
 ולכן ניתן לומר שהנגזרת החלקית לפי w היא חיובית, והנגזרת החלקית לפי b היא שלילית.