

# תרגיל בית מספר 1

**נושא: שפת C++ כשפת C משופרת**

**תאריך הגשה: יום ד', 23/11/2022, בשעה 23:59**

**הגשה ביחידים**

**בהצלחה רבה!**

## תיאור התרגיל

מטרתו של תרגיל בית זה הינה כפולה – הראשונה, יישור קו מבחינת רמת התכנות הנדרשת בתחילת הקורס; השנייה, תכנות בשפת C++ כשפת C משופרת, אך לאו דווקא כשפה מונחית עצמים בשלב זה. התרגיל יעסוק בבניית מחשבון סטטיסטי עבור נתונים המתקבלים מהמשתמש, אשר ישלב רכיבי קלט/פלט, חישוב מתמטי בסיסי, וניהול זיכרון.

## מחשבון סטטיסטי: אפיון

בבסיס העבודה עומדת הזנת נתונים גולמיים, המיוצגים באמצעות וקטורים ממשיים ממימד קבוע, אל הזיכרון, ושאלות הדורשות חישוב סטטיסטי עליהם.

נגדיר **תצפית** כוקטור ממשי המייצג אוסף של נתונים גולמיים. **מימד התצפית** הוא מספר הקואורדינטות המתארות אותה. עליכם לתכנן ולממש מחשבון סטטיסטי אשר מטרתו לקלוט באופן מתמיד תצפיות מן המשתמש **ממימד נתון וקבוע** ולתמוך בפעולות הבסיסיות הבאות:

- [1] הזנת תצפית אל הזיכרון תחת שם נתון "New observation"
- [2] שליפת תצפית מן הזיכרון ע"פ שם והדפסתה בערוץ הפלט הסטנדרטי "Print observation"
- [3] הדפסת וקטור התוחלת של כלל התצפיות "Expected value vector"
- [4] הדפסת מטריצת השונות המשותפת של כלל התצפיות "Covariance matrix"
- [5] יציאה "Exit"

עם הרצת התכנית, המערכת תציג למשתמש תפריט פעולות התואם את הפעולות הבסיסיות שהוצגו, ותמתין לבחירתו; לאחר ביצוע פעולות [1] עד [4] המערכת תחזור למצב "המתנה" לבחירה הבאה, ורק פעולה [5] תביא לסיומה של

התכנית. הזנה של מספר פעולה שאינו קיים (שונה מ [1-5]) תגרור הדפסת הודעת שגיאה וחזרה למצב המתנה)

Invalid option.\n

- התקשורת מול המשתמש תתנהל בערוצי הקלט והפלט הסטנדרטיים, כאשר הודעות שגיאה תצאנה אל ערוץ פלט השגיאה.

## הזנת תצפית לזיכרון המחשבון

בבחירת פעולה זו [1], התכנית תקלוט מן המשתמש תצפית ותשמור אותה במבנה הנתונים הפנימי שלה תחת שם שיימסר ע"י המשתמש. קליטת תצפית מן המשתמש תתנהל לפי הפרוטוקול הבא, ואמורה להתלוות בטקסט שינחה את המשתמש:

- א. בקשה מהמשתמש להזנת שם התצפית "Enter observation name: "
- ב. הקלדת שם התצפית (זהו המזהה הייחודי של כל תצפית, המהווה מחרוזת כלשהי בעלת לכל היותר 16 תווים) ע"י המשתמש והקלדת Enter, ניתן להניח כי שם התצפית תקין.
- ג. בקשה מהמשתמש להזנת ערכי התצפית "Enter observation values: "
- ד. הזנת ערכי התצפית (מספרים ממשיים מטיפוס double) ע"י המשתמש, כאשר כל ערך יופרד באמצעות רווח.
- ה. עדכון מבנה הנתונים הפנימי של המחשבון וחזרה אל התפריט הראשי.

## תכנות בשפת C++, סתיו 2022-23

במקרה שתצפית בעלת שם זהה כבר קיימת בזיכרון המחשבון, עדכון מאוחר ידרוס לחלוטין כל מידע קודם שהוזן למחשבון. במקרה של חריגה מפרוטוקול זה מצד המשתמש, על המערכת להוציא הודעת שגיאה ולחזור לתפריט הראשי.

Invalid observation.\n

### שליפת תצפית מזיכרון המחשבון והדפסה

בבחירת פעולה זו [2], על התכנית להדפיס תצפית יחידה ששמה נמסר ע"י המשתמש.

הפרוטוקול כולל פניה טקסטואלית אל המשתמש בבקשה להקלדת שם התצפית להדפסה. "Enter observation name: "

בהינתן שם חוקי אשר קיים בזיכרון, תודפס התצפית בפורמט שיתואר בהמשך. בהינתן שם שאינו קיים בזיכרון, על המערכת להוציא הודעת שגיאה ולחזור לתפריט הראשי.

Invalid or nonexistent observation.\n

### חישוב וקטור התוחלת של התצפיות הקיימות בזיכרון

בבחירת פעולה זו [3], התכנית תחשב את וקטור התוחלת של כל התצפיות הקיימות בזיכרון ותדפיס אותו אל ערוץ הפלט הסטנדרטי בפורמט שיתואר בהמשך.

וקטור התוחלת של התצפיות מוגדר כוקטור במימד התצפית אשר כל קואורדינטה שלו מהווה את ממוצע התצפיות באותה קואורדינטה (ראו פירוט בנספח).

במקרה שזיכרון המחשבון ריק בשלב של בחירת פעולה זו, על המערכת להוציא הודעה שגיאה ולחזור לתפריט הראשי. בכל מקרה אחר, המחשבון יחזור לתפריט הראשי לאחר החישוב וההדפסה.

Empty calculator.\n

### חישוב מטריצת השונות של התצפיות הקיימות בזיכרון

בבחירת פעולות זו [4], התכנית תחשב את מטריצת השונות המשותפת של כל התצפיות הקיימות בזיכרון ותדפיס אותה אל ערוץ הפלט הסטנדרטי. מטריצת השונות המשותפת של התצפיות הינה מטריצה ריבועית במימד התצפית אשר כל אלמנט שלה מכמת מדד קשר בין קואורדינטות מסוימות עבור כלל התצפיות. הסבר מפורט ומתכון לחישוב המטריצה מופיעים בנספח.

במקרה שזיכרון המחשבון ריק בשלב של בחירת פעולה זו, על המערכת להוציא הודעה שגיאה ולחזור לתפריט הראשי. בכל מקרה אחר, המחשבון יחזור לתפריט הראשי לאחר החישוב וההדפסה.

Empty calculator.\n

**פורמט הדפסת וקטור:** השורה הראשונה תכיל את שם התצפית (או במקרה של חישוב וקטור התוחלת את המחרוזת mean), רווח, סימן שווה, רווח וסוגר מרובע שמאלי – ולאחר מכן את כל הקואורדינטות של הוקטור – מופרדות באמצעות רווח – ולסיום, סוגר מרובע ימני. לדוגמא, כך תיראה הדפסה של וקטור התוחלת (פעולה [3]) עבור תצפיות ממימד 5:

mean = [ -0.43 0.51 0 0 0 ]

**פורמט הדפסת מטריצת השונות המשותפת:** השורה הראשונה תכיל את המחרוזת cov, סימן שווה, וסוגר מרובע שמאלי – מופרדים באמצעות רווח.

כל אחת מן השורות הבאות תכיל שורת אלמנטים של המטריצה, כאשר כל אלמנט יופרד מאחר באמצעות רווח בודד. השורה האחרונה בהדפסה, שתבוא לאחר השורה האחרונה של המטריצה, תכיל רק סוגר מרובע ימני. לדוגמא, מטריצת שונות משותפת ממימד 6 תודפס כך לפלט הסטנדרטי:

## תכנות בשפת C++, סתיו 2022-23

```
cov = [  
1 0.5 0 0 0 0  
0.5 1 0 0 0 0  
0 0 1 0 0.7 0  
0 0 0 1 0 0  
0 0 0.7 0 1 0  
0 0 0 0 0 1  
]
```

### הרצת התכנית

בעת הרצת התכנית יימסרו לה שני ארגומנטים: הראשון יציין את מימד התצפיות והשני יציין את מספר התצפיות המקסימלי שעל המחשבון להכיל. לדוגמא, לאחר קומפילציה של התכנית לכדי קובץ הרצה בשם `statObservations`, ההרצה הבאה בטרמינל תאתחל מחשבון המיועד לקליטת תצפיות ממימד 10 ולכל היותר 1500 במספר:

```
>> ./statObservations 10 1500
```

כל ניסיון להריץ את התכנית אחרת, או הרצה עם פרמטרים שגויים, תביא להוצאת הודעת שגיאה ופירוט הדרך להרצת התכנית.

```
Invalid arguments <int> <int>.\n
```

### הנחות עבודה ופורמט

- על מבנה הנתונים בו מאוחסנות התצפיות להיות מוקצה דינאמית על ה-Heap. יש לכם חופש מוחלט בבחירת טיפוס מבנה הנתונים.
- יש להשתמש בטיפוס נתונים `float` בעת חישוב היחסים; בעת ההדפסה, יש לעשות שימוש בברירת המחדל של אופרטור `<<` (אין צורך לפרמט אחרת).
- שם תצפית הינו מחרוזת ייחודית שאינה עולה באורכה על 16 תווים כלשהם.
- מספר התצפיות שניתן לאחסן בזיכרון חסום ע"י הארגומנט השני שניתן לתכנית בעת הרצתה; ניסיון להזין תצפיות מעבר לחסם זה, יגרור הודעת שגיאה המציינת כי מספר התצפיות המקסימאלי כבר מאוחסן בזיכרון.

### דגשים

- יש לתכנן מראש את מבנה התכנית, ולהגדיר בהתאם את המרכיבים איתם תעבדו.
- יש לממש את עבודת הקלט והפלט באמצעות רכיבי C++ ולא באמצעות פונקציות C.
- הקצאה דינאמית צריכה להתבצע רק באמצעות הפקודות `new` ו-`delete`.
- עבודה עם מחרוזות יכולה להתבצע באמצעות טיפוס `std::string`, אך אין דרישה לכך.
- **בתרגיל בית זה אין להשתמש בספריית STL; מבני הנתונים והאלגוריתמים צריכים להיות ממומשים על-ידיכם.**
- יש לבדוק תקינות קלטים ולהציג הודעות שגיאה מתאימות.
- עליכם לוודא כי התכנית עוברת קומפילציית g++ התואמת את הקומפיילר שעל שרת החוג ללא כל שגיאות או אזהרות כלשהן, ורצה בהצלחה.
- עליכם לתעד את הקוד באמצעות הערות המתארות בקצרה את המודולים והפונקציות השונים.
- יש להריץ את הבדוק האוטומטי על שרת החוג בטרם ההגשה בכדי לוודא תאימות ונכונות של ההגשה: **hwcheck** (או לחילופין עשו שימוש בפרוטוקול HTML בקישור [cswb.telhai.ac.il](http://cswb.telhai.ac.il)).

- עליכם להגיש במערכת Moodle קובץ ארכיב מטיפוס zip בלבד, ששמו כולל את קוד הקורס ('44'), שם התרגיל ('ex1') ותעודת הזהות של הסטודנט/ית המגיש/ה, מופרדים בקו תחתי בפורמט הבא: **44\_ex1\_studID.zip**.
- על ארכיב zip זה להכיל את כל קבצי המקור (ממשק/מימוש) הנדרשים לקומפילציה, והוא רשאי להכיל תיעוד טקסטואלי; מבחינת טיפוס קבצים, עליו לכלול רק קבצים עם סיומות **\*.cpp \*.h \*.txt**.
- לדוגמה: על סטודנט שמספר הזיהוי שלו הינו 012345678 להגיש ארכיב בשם 44\_ex1\_012345678.zip הכולל את כל קבצי המקור של הפרוייקט, ללא תיקיות כלשהן, ורשאי להכיל קובץ טקסטואלי לתיעוד.

אי-הקפדה על ההנחיות, כולל פורמט ההגשה הדיגיטלי, תגרור הורדה בציון התרגיל.

לא תתקבלנה הגשות באיחור!

בהינתן  $M$  תצפיות מממד  $n$ ,  $\vec{x}^{(1)}, \vec{x}^{(2)}, \dots, \vec{x}^{(M)} \in \mathbb{R}^n$ , התוחלת של קבוצת התצפיות הללו היא וקטור ממשי מממד  $n$  בעל הצורה המפורשת הבאה:

$$\vec{\mu} = \begin{pmatrix} \frac{1}{M} \cdot \sum_{i=1}^M x_1^{(i)} \\ \frac{1}{M} \cdot \sum_{i=1}^M x_2^{(i)} \\ \vdots \\ \frac{1}{M} \cdot \sum_{i=1}^M x_n^{(i)} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix}$$

כלומר, וקטור בו כל קואורדינטה מהווה את הממוצע של הקואורדינטות המקבילות עבור כל  $M$  התצפיות.

מטריצת השונות המשותפת הינה מטריצה ריבועית מעל הממשיים מממד  $n$  אשר אלמנט  $ij$  שלה מתייחס לקשר בין קואורדינטה  $i$  לבין קואורדינטה  $j$  עבור כלל התצפיות:

$$\mathbf{C} = (c_{ij}) :$$

$$\text{שימו לב שהנירמול} \left\{ \begin{array}{ll} c_{ij} = \frac{1}{M} \cdot \sum_{k=1}^M (x_i^{(k)} - \mu_i) \cdot (x_j^{(k)} - \mu_j) & M = 1 \\ c_{ij} = \frac{1}{M-1} \cdot \sum_{k=1}^M (x_i^{(k)} - \mu_i) \cdot (x_j^{(k)} - \mu_j) & M > 1 \end{array} \right.$$

משתנה בין מקרה בו יש תצפית בודדת ( $M=1$ ) לבין מקרה של תצפיות מרובות.