

אין צורך לבצע אימות קלט.

1 שימוש במדדים

1. הורידו את הקובץ `am_lab8.py` מהמודל והשלימו אותו לפי ההנחיות.
2. הורידו את הקובץ `data.tsv` מהמודל. הקובץ מכיל `cross sample differential coverage` בין רצפים גנטיים חלקיים. בשיטה זו מחשבים את הכיסוי הממוצע של כל רצף במספר דוגמאות שונות. נצפה כי רצפים חלקיים השייכים לאותו הגנום יקבלו ערכים דומים, והוקטור שלהם יהיה יותר דומה.
3. כתבו פונקציה הטוענת את המידע למילון. כל שורה מייצגת רצף אחד. מפתחות המילון יהיו שמות הרצפים, ערכי המילון יהיו מערכי `numpy` של הערכים.
- אפשר להעזר במתודה `readline` כדי לקרוא שורה אחת מקובץ פתוח. אפשר לפצל שורה עם המתודה `split`.
4. ממשו את המדדים הבאים, כל אחד בפונקציה בפני משלו, על הפונקציה לקבל שני משתנים, וקטורים של `numpy`:

l_p :

$$d_p(x, y) = \left(\sum_{i=1}^n |x_i - y_i|^p \right)^{\frac{1}{p}}$$

כמופיע בעמוד 6 של מצגת 2.

בחרו ערך p כלשהו, וממשו פונקציה המקבלת רק x ו- y הקוראת לפונקציה `lp` עם ערך קבוע. זאת אומרת, אם בחרתי ב- $p = 9$ אוסיף פונקציה באופן הבא:

```
1 def l9(x, y):
2     return lp(x, y, 3)
```

בחרו בערך שאינו תשע. אפשר להוסיף מספר ערכים שונים (ופונקציה לכל ערך).

1 - Cosine Similarity:

$$1 - \frac{\sum_{i=1}^n x_i \cdot y_i}{\sqrt{\sum_{i=1}^n x_i^2} \cdot \sqrt{\sum_{i=1}^n y_i^2}}$$

1 - Weighted Jaccard Distance:

$$1 - \frac{\sum_{i=1}^n \min(x_i, y_i)}{\sum_{i=1}^n \max(x_i, y_i)}$$

יש לממש את הפונקציות בעצמכם!

5. ממשו פונקציה היודעת לקבל את מילון הדוגמאות, ולהחזיר מילון של מטריצות ערכי מדדים. במילון המפתחות יהיו שמות פונקציית הממד בתור מחרוזת, הערכים יהיו מטריצה של ערכי הממד עבור כל זוג מערכים מהקלט. לדוגמה, התא במקום 0,0 יהיה ערך הממד של רצף גנטי מול עצמו, במקום 0,1 מול רצף אחר וכך הלאה.
- כדאי, אך לא חובה, להוסיף פונקציית עזר, הפונקציה תקבל את כל המערכים עבורם יש לחשב את המרחק ואת פונקציית המרחק בתור משתנה. תיצור את המטריצה ותמלא אותה בעזרת לולאה מקוננת.
6. הוסיפו פונקציה היודעת לקבל את מילון מטריצות ערכי הממד, רשימה של שמות הרצפים (הם המפתחות במילון המערכים שטענו בהתחלה) ושם קובץ לשמירת האיור, הפונקציה תייצר איור עם תרשים עבור כל מדד שהפעלנו. תוכלו להעזר במתודה `subplots`.
- הציגו את ערכי הממד בתור תמונה של רמות אפור (`imshow` עם `cmap='gray'`), שימו את שם הממד בתור כותרת (`set_title`), את שמות הרצפים בתור האינדקסים בתמונה בעזרת `ax[i].set_xticks(range(len(labels)), labels)` `ax[i].set_yticks(range(len(labels)), labels)` עבור `labels` שמות הרצפים (החליפו שמות משתנים בהתאם).
- לאחר מכן האיור ישמר בשם שנבחר.
- שימו לב : בתמונה של רמות אפור, ערך נמוך יותר יופיע כהה יותר וערך גבוה יותר יהיה בהיר יותר.
7. חברו הכל ביחד בפונקציה המקבלת את קובץ הקלט, שם הפלט לאיור ומייצרת אותו.

2 תרגיל הגשה עם ציון

דברים שיובילו להורדת ציון:

פורמט הגשה	השפעה על ציון
קובץ זיפ	-50
יותר מקובץ קוד אחד	-5
קוד מחוץ לקובץ py	-15
קוד ללא הזחות	-30
קוד שלא רץ	-40
קוד לא מסודר	$-d \in \{n n \in \mathbb{N} \wedge n \geq 5 \wedge n \leq 20\}$
העתקה	אפס במעבדה אחת לפחות

מה יש להגיש: קובץ פייתון אחד בלבד העונה על כל הסעיפים עבור עבודה עם שורשי מספרים מרוכבים.