**כל מה שקרה לפני**

החלטנו שאני אעשה את המרחק הכימי-פיזיקלי בין רצפים.

בהתחלה עשיתי משהו עם מוטציה אחת שעבד מעולה: ייצרתי את כל האפשרויות להחלפות ומיינתי לפי סדר עולה של עלות ההחלפה, מתוך הרשימה הנתונה של הרצפים על כל רצף ניסיתי להחליף את המוטציות בסדר העולה ולחפש האם הרצף המוטנטי קיים ברשימה, ברגע שמצאתי רצף מוטנטי הוספתי את הרצף המקורי ואת הרצף המוטנטי לtuple והמשכתי לרצף הבא ברשימה.

אחרי זה ייצגתי את הכל בגרף והיה יפה ועבד מהר.

סול אמר שהוא היה רוצה לדעת האם עדיף לעשות שתי מוטציות או יותר על פני מוטציה אחת.

התחלתי לעשות את זה בשיטה המקורית, אבל נתקעתי בחלק של לייצר את כל האפשרויות להחלפה, לוקח מלא זמן.

סול הציע את הפתרון הבא:

נייצר פונקציית האש ונעשה האש על כל הרצפים הקיימים.

על כל רצף נייצר את כל הרצפים שיכולים להתקבל כתוצאה ממוטציה אחת, שתיים שלושה וכו' ונעשה עליהם גם כן האש.

נאחד בין הרשימות, נמיין את הרשימה החדשה ונראה האם יש רצפים כפולים.

**3.3**

סיימתי עם הפתרון של סול, אבל זה לוקח המון זמן, נראה לי שהמיון גורם לזה להיות מאוד כבד.

**4.3**

היום הייתה לי שיחה עם סול והוא הציע לי פתרון לבעיה של המהירות.

מה שעשיתי עד עכשיו היה לייצר לכל רצף מוטציות, לעשות לכולן האש, למיין ולבדוק האם יש כפילויות.

הוא אמר שבגלל שיש פה המון מיונים זה יכול לגרום לאיטיות, ולכן הוא הציע לכל רצף לייצר את המוטציות שלו, להכניס את הכל לטבלה גדולה ששומרת את המוטציות ואת המקור ולעשות אותו דבר לרצף הבא בלי למיין עדיין.

אחרי שסיימתי לייצר את כל המוטציות לכל הרצפים, אני אעשה האש לכולם ואז אעשה מיון אחד לכל המוטציות.

אחרי שעשיתי את המיון אני אחפש כאלה שחוזרים על עצמם ואדע למי הם שייכים לפי המקור.

**5.3**

היום התחלתי לעבוד על הפתרון שסול נתן לי, צורת המימוש:

מייצרת רשימה שתהפוך לטבלה שתיראה כך:

שלוש עמודות – mut, mut hash, org.

בעמודת הmut יהיה לנו מוטציות של רצף

בעמודת הmut hash יהיה לנו את ההאש של הרצף

בעמודת הorg יהיה לנו את הרצף המקורי

כלומר, מייצרת רשימה של רשימות באורך שלוש, מאוחר יותר אהפוך את הרשימה לטבלה עם pandas.

בהתחלה אני עוברת על כל הרצפים המקוריים ומוסיפה אותם לעמודת הmut, בעמודת הorg אני שמה סטרינג ריק.

אחרי זה אני עוברת על כל רצף, מייצרת לו מוטציות ומכניסה לטבלה כך שבעמודת הmut יש את הרצף המוטנטי ובעמודת הorg הרצף המקורי.

אחרי שסיימתי לייצר את כל המוטציות אני עוברת על כל המוטציות ועושה להן האש.

בשלב הזה אני הופכת את הרשימה לטבלה של pandas, הסיבה שאני עושה את זה ככה ולא מייצרת טבלה מראש ומכניסה אליה כל פעם ערכים היא שעלות ההכנסה לרשימה היא O(1) בעוד שעלות ההכנסה לטבלה בpandas היא O(n) מכוון שהוא מייצר עותק חדש של הטבלה.

עכשיו אני ממיינת את הטבלה לפי עמודת ההאש ומפלטרת את הטבלה לפי הרצפים המקוריים, כל פעם לפי רצף אחד, ומחפשת כפילים.