

N1)

a) კოდი არის problem1/gibbs.py ფაილში

b) მოთხოვლი ინფორმაცია არის problem1/PWM ფაილებში, data-ს მიხედვით.

c) სურათები და დაგენერირების ინსტრუქცია არის problem1/problem1_c ფოლდერში

N2)

a) კოდი არის problem2/problem2.py ფაილში.

გაშვების ბრძანება: `python problem2.py allinter allintercons yeast_motifs.txt`

დასაბრუნებელი ტექსტური ფაილი არის A.txt

b) მიკერძოებას რასაც ვამჩნევ არის, რომ ხშირ kmer-ებში საკმაოდ ხშირად გვხვდება A და T ნუკლეოტიდები. კონსერვირებულ რეგიონებს დიდად ვერ ვამჩნევ მიკერძოებას, შეიძლება ოდნავ უფრო GC rich იყოს.

Which of the two lists should we use to direct further inquiry into yeast transcription factor binding sites?

კონსერვირებული რეგიონები უნდა გამოვიყენოთ, რადგანაც ხშირად შემხვედრი რეგიონები შეიძლება უბრალოდ ხშირად გვხვდება და არაფერს ნიშნავს. სამაგიეროდ, რეგიონი რომ კონსერვირებულია, უფრო მაღალია ალბათობა, რომ რაღაც ფუნქცია ჰქონდეს და შემთხვევით არ იყოს შემონახული.

Which known motifs does your scan of 6-mers find?

kmer-ების yeast_motifs.txt ფაილში ძებნის მიღებული შედეგი წერია B.txt ფაილში.

N3)

a) კოდი არის problem3/nussinov.py ფაილში. გაშვების ბრძანება: `python nussinov.py`

b) 43.992

c) 50-ზე არის 21.065, 25-ზე 9.888, 75-ზე 32.54. ანუ, ფაქტობრივად წრფივთან საკმაოდ მიახლოებული დამოკიდებულებაა.

d) სიმეტრიულად თუ შევცვლით ალბათობებს, ანუ G და C ს მოსვლის ალბათობებს თუ მიახლოებით თანაბარს ავიღებთ და ასევე A და U ალბათობებსაც ერთმანეთთან ახლოს მყოფს, მაგ შემთხვევაში ქულაც დაახლოებით იგივე რჩება, ოდნავ დამოკიდებულია GC კონტენტზე, ანუ ზრდით იზრდება და შემცირებით მცირდება, მაგრამ მნიშვნელოვნად არა. თუ ასიმეტრიულად ავიღებთ ალბათობებს, მაგალითად 0.8 რომ იყოს C ს ალბათობა და 0.1 G-ს, ასეთ შემთხვევაში კი მკვეთრად იკლებს ქულა. ყველაფერი დამოკიდებულია იმაზე თუ

რამდენად “კარგად” შეუძლია წყვილის პოვნა თითოეულ ნუკლეოტიდს და კომპლემენტარული ნუკლეოტიდების სიმეტრიულ განაწილებებზე ეს ყველაზე კარგად ხდება.

e) სიგრძე რადგანაც ვიცით, ნუსინოვის ქულის მიახლოებითი წინასწარმეტყველება შეგვიძლია განაწილებების შესწავლით. მიახლოებების გაუმჯობესება შეგვიძლია AU და GC content-ზე უკეთესი დაკვირვებით და ზოგადად განაწილებებზე დაკვირვებით, რომ დავადგინოთ სქორთან დამოკიდებულება. ამ დამოკიდებულების შესწავლით ალბათ ML მოდელიც კი შეგვიძლია ავაგოთ, რომელიც ნუსინოვის ქულის პრედიქციას გააკეთებს. და ისევე შემთხვევებში, როდესაც N^3 time complexity ნელია წესით პრედიქციის ალტერნატივა უფრო სწრაფი უნდა იყოს, და შეიძლება საკმარისად კარგ პასუხს გვაძლევდეს, რომ თვითონ ალგორითმის გაშვება აღარ ღირდეს.