



دانشگاه صنعتی امیرکبیر
(پلی تکنیک تهران)

بسمه تعالی

تمرین سری چهارم بیوانفورماتیک



دانشکده مهندسی

نکته مهم: تمرین چهارم در سایت کوئرا قرار داده شده است و شما می‌توانند از طریق [این لینک](#) در کلاس مربوطه در کوئرا با کلید Bio1400 ثبت نام نمایید. تحویل پاسخ‌های تشریحی نیز از طریق همین صفحه‌ی کوئرا انجام خواهد شد.

سوال ۱)

در مورد PSI Blast به سوالات زیر پاسخ دهید:

الف) چگونه می‌توان از روی تنها یک کوئری که دنباله پروتئینی است، PSSM را ساخت؟

ب) profile drift چیست و چه تاثیری در نتیجه PSI Blast دارد؟

ج) در صورت کاهش آستانه (ترشولد) مورد استفاده در قسمت جست‌وجوی این الگوریتم، مشکل بیان شده در قسمت قبل چه تغییری می‌کند؟

سوال ۲)

الف) برای دنباله‌های داده شده، ماتریس PSSM را محاسبه کنید. مقدار pseudocount را برابر صفر و مقدار احتمال background را برابر مقدار $\text{random chance} = 0.25$ در نظر بگیرید. (اعداد را تنها تا ۳ رقم اعشار حفظ کنید)

توجه: مانند مثال‌های اسلایدهای درس مراحل رسیدن به پاسخ را بنویسید.

ATGCCG
AAGATT
TACTCA
CTGAGG
CACCTG

ب) اگر بخواهیم logo را محاسبه کنیم، اعداد مربوط به هر residue را تنها برای جایگاه اول دنباله‌ها (اولین ستون) محاسبه کنید.

سوال ۳)

الف) برای دنباله‌های زیر profile را محاسبه کنید. مقدار pseudocount را برابر یک در نظر بگیرید و مقدار احتمال background را از روی تعداد هر residue در دنباله‌های داده شده محاسبه کنید. (اعداد را تنها تا ۳ رقم اعشار حفظ کنید)

توجه: مانند مثال‌های اسلایدهای درس مراحل رسیدن به پاسخ را بنویسید.

AT - G - CCG
AA - G - CTT
T - ACT - CA
CTGACGGA

(ب) احتمال دنباله AA-CTCTG چند برابر حالت تصادفی است؟

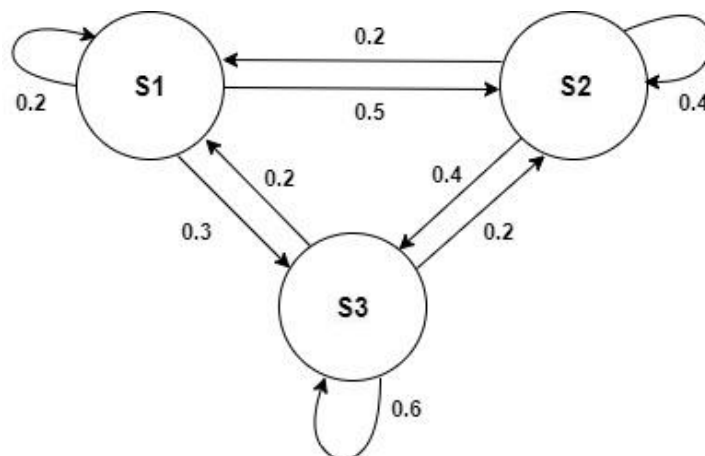
(ج) محتمل‌ترین دنباله را به دست آورید.

سوال ۴

با توجه به مدل مخفی مارکوف زیر و پارامترهای داده شده دو سوال زیر را پاسخ دهید. (همان‌طور که مشخص شده، در π اعداد را به ترتیب برای S1 و S2 و S3 و برای b اعداد را به ترتیب برای A, C, G و T در نظر بگیرید)

$$\pi = \begin{bmatrix} \pi_{s1} \\ \pi_{s2} \\ \pi_{s3} \end{bmatrix}, \quad \pi = \begin{bmatrix} 0.7 \\ 0.2 \\ 0.1 \end{bmatrix}$$

$$b_{si} = \begin{bmatrix} b_{si}(A) \\ b_{si}(C) \\ b_{si}(G) \\ b_{si}(T) \end{bmatrix}, \quad b_{s1} = \begin{bmatrix} 0.2 \\ 0.3 \\ 0.2 \\ 0.3 \end{bmatrix}, \quad b_{s2} = \begin{bmatrix} 0.1 \\ 0.1 \\ 0.4 \\ 0.4 \end{bmatrix}, \quad b_{s3} = \begin{bmatrix} 0.4 \\ 0.2 \\ 0.1 \\ 0.3 \end{bmatrix}$$



(الف) ماتریس زیر را که همان ماتریس α است تکمیل کرده و احتمال دنباله ATCG را بیابید.

	A	T	C	G
S1				
S2				
S3				

(ب) ماتریس زیر که همان ماتریس δ است را تکمیل کرده و محتمل‌ترین دنباله حالت‌ها را برای دنباله CGAT به دست آورید.

	C	G	A	T
S1				
S2				
S3				

نکات تحویل تمرین

- (۱) تمرینات را به صورت انفرادی انجام دهید.
- (۲) پاسخ‌های خود را با کیفیت مناسب و خوانا اسکن کرده و یا تایپ شده به صورت یک فایل درآورید و با نام "HW4_studentNumber_FirstnameLastname.pdf" در سایت بارگذاری نمایید.
- (۳) در صورت پیداشدن هرگونه کپی چه در سوالهای تشریحی و چه برنامه‌نویسی نمره‌ی هر دو نفر ۱۰۰- در نظر گرفته خواهد شد. این نکته را جدی بگیرید تا بعداً به مشکل نخورید.
- (۴) اشکالات خود در رابطه با بخش تشریحی تمرین (همین سوالات) را می‌توانید از طریق این ایمیل‌ها بپرسید.
majdiad09@gmail.com
همچنین اشکالات در رابطه با کدنویسی را نیز می‌توانید با ایمیل زیر مطرح کنید.
mmnafar57@gmail.com
- (۵) مهلت ارسال این تمرین تا پنج‌شنبه ۱۹ خرداد است و پس از این تاریخ حداکثر تا دو روز به ازای هر ساعت تاخیر ۱ درصد جریمه در نظر گرفته خواهد شد. بعد از این مدت پاسخ سوالات در کانال درس ارسال شده و دیگر نمره‌ای به تمرین تعلق نخواهد گرفت.