

مبانی بیوانفورماتیک

نیم سال اول ۰۳-۰۲

گردآورندگان: عارف شه بخش، علی رشیدی، کوثر یوسفی

دانشکده‌ی مهندسی کامپیوتر

آزمایش فرضیه، آزمون adjustment / P-value، تست فیشر، تحلیل مسیر، مدل های مارکوف، موتیف یابی

تمرین سری چهارم

زمان تحویل: ۱۵ دی

توضیحات: پاسخ سوالهای ۱ تا ۵ را پس از تایپ یا تصویر از دستنویس در یک فایل zip به صورت Stname_ST#_HW4 در قسمت مربوط به تمرین کوثرای درس آپلود کنید. و تمرینهای کدنویسی مربوط به سوال ۶ در Nootbook مربوطه قرار داده شده است. همچنین تمرینهای کدنویسی مربوطه در سایت Rosalind قرار داده شده است.

مسئله ۱: (۱۰ امتیاز)

یک تیم تحقیقاتی بیوانفورماتیک در حال مطالعه اثر یک داروی جدید بر بیان ژن هستند. این تیم داده های بیان ژن را از دو گروه سلول جمع آوری کرده است: یکی تحت درمان با دارو (گروه A) و دیگری درمان نشده (گروه کنترل، گروه B). هر گروه دارای ۱۰۰ سلول است و سطح بیان یک ژن خاص مورد نظر در هر سلول اندازه گیری شده است. این تیم می خواهد بداند که آیا داروی جدید تأثیر قابل توجهی بر سطح بیان ژن مورد نظر دارد یا خیر. آنها فرض می کنند که میانگین سطح بیان ژن در سلول های تیمار شده (گروه A) با سلول های تیمار نشده (گروه B) متفاوت است. از شما خواسته می شود برای ارزیابی این ادعا یک آزمون فرضیه انجام دهید. توضیح دهید که چگونه به این کار نزدیک خواهید شد و چه نتایجی می توانید از نتایج ممکن مختلف بگیرید.

مسئله ۲: (۲۰ امتیاز)

فرض کنید در حال انجام یک تراز چند دنباله با استفاده از یک جفت مدل مارکوف پنهان (جفت-HMM) هستید. شما دو دنباله DNA به نام های x و z دارید و می خواهید بهترین تراز بین آنها را پیدا کنید. دنباله ها به شرح زیر است:

$x = \text{TTCCG}$

$z = \text{CCGTT}$

فرض کنید دنباله حالت زیرین $y = |x| |z|$ است. در این توالی حالت، I یک حالت درج، A نشان دهنده وضعیت تطبیق، و x و z دو توالی DNA هستند.

احتمال انتشار را به صورت زیر فرض کنید:

احتمالات انتشار برای حالت I : $P(A | I) = 0.1, P(T | I) = 0.2, P(C | I) = 0.3, P(G | I) = 0.4$

احتمالات انتشار برای حالت A : $P(A | A) = 0.4, P(T | A) = 0.3, P(C | A) = 0.2, P(G | A) = 0.1$

۱. با استفاده از جفت-HMM، احتمال تراز داده شده را محاسبه کنید.

۲. اگر تراز نادرست باشد، احتمال تراز صحیح چقدر است؟

مسئله ۳: (۲۰ امتیاز)

مطالعه ای را در نظر بگیرید که در آن شما در حال بررسی ارتباط بین استفاده از دو درمان مختلف (داروی X و داروی Y) و احتمال درمان سرطان پوست هستید. شما داده های زیر را جمع آوری کرده اید:

	مجموع	دارو Y	دارو X
درمان شده	۱۳۰	۳۰	۱۰۰
درمان نشده	۲۵۰	۱۰۰	۱۵۰
مجموع	۳۸۰	۱۳۰	۲۵۰

فرضیه صفر شما (H_0) این است که هیچ ارتباطی بین استفاده از دو روش درمانی و احتمال درمان سرطان پوست وجود ندارد.

فرضیه جایگزین شما (H_A) این است که بین استفاده از این دو درمان و احتمال درمان سرطان پوست ارتباط وجود دارد.

۱. نسبت شانس را با استفاده از داده ها محاسبه کنید.

۲. نسبت شانس را در چارچوب فرضیه ها تفسیر کنید.

۳. مقدار p را برای آزمون دقیق فیشر محاسبه کنید.

۴. p -value را در چارچوب فرضیه ها تفسیر کنید.

مسئله ۴: (۱۵ امتیاز)

مدل های پنهان مارکوف (HMMs) به طور گسترده در بیوانفورماتیک برای کارهایی مانند پیش بینی ژن، هم ترازی توالی و پیش بینی ساختار پروتئین استفاده می شوند.

یک مدل پیش بینی ژن با استفاده از HMM مرتبه اول با پارامترهای زیر طراحی شده است:

• جدول انتقال حالات

state	Start	Exon	Intron	Stop
Start	۰.۳	۰.۴	۰.۳	۰
Exon	۰	۰.۷	۰.۲	۰.۱
Intron	۰	۰.۱۵	۰.۸	۰.۰۵
Stop	۰	۰	۰	۱

• جدول انتشار

state\emission	A	C	G	T
Start	۰.۲۵	۰.۲۵	۰.۲۵	۰.۲۵

Exon	۰.۳۵	۰.۱۵	۰.۱۵	۰.۳۵
Intron	۰.۱	۰.۴	۰.۴	۰.۱
Stop	۰.۲۵	۰.۲۵	۰.۲۵	۰.۲۵

- الف) احتمال تولید دنباله "ACGTA" از مدل HMM، با شروع از **Start** چقدر است؟
 ب) اگر دنباله "ACGGT" را مشاهده کنید، محتمل ترین حرف بعدی این توالی چه خواهد بود؟

مسئله ۵: (۲۰ امتیاز)

مدل مارکوف یک مدل آماری است که انتقال حالت یک سیستم را بین سایر حالات ممکن مدل میکند، با این فرض که حالت بعدی سیستم فقط به حالت فعلی سیستم بستگی دارد بدین معنی که دانستن حالت فعلی ... حالت بعدی را از حالات گذشته مستقل میکند. در بیوانفورماتیک، مدل های مارکوف برای کاربردهای مختلف از جمله تحلیل توالی استفاده می شوند. یک توالی **DNA** متشکل از نوکلئوتیدهای **A, C, G** و **T** را در نظر بگیرید. یک مدل مارکوف مرتبه اول با احتمال های انتقال زیر به شما داده شده:

state	a	c	g	t
a	۰.۳	۰.۲	۰.۲	۰.۳
c	۰.۲۵	۰.۳	۰.۲	۰.۲۵
g	۰.۴	۰.۱	۰.۲	۰.۳
t	۰.۲	۰.۴	۰.۳	۰.۱

- الف) مطلوب است احتمال تولید دنباله "AGCT" با شروع از حالت **A** منشا شده از این مدل مارکوف.
 ب) محتمل ترین تغییر حالت پس از مشاهده دنباله "ACGGT" طبق این مدل مارکوف چیست؟
 ج) گراف احتمالاتی حالت گذار این مدل را ترسیم کنید.
 د) وجود یا عدم وجود توزیع استشنری این مدل را ثابت کنید.

مسئله ۶: (۱۵ امتیاز)

برای حل این سوال به **Nootbook** مربوطه مراجعه کنید (در نهایت فایل نهایی **Nootbook** کامل شده را به عنوان پاسخ سوال ارسال کنید)

