



ژنومیک محاسباتی

مطهری و فروغمند
پاییز ۱۴۰۰

امتحان میان ترم

پاسخ سوال های امتحان میان ترم
نگارنده: الهه بدلی - ۹۸۲۰۹۰۷۲

۱ پاسخ سوال: پاسخ کوتاه ۱

الف) یعنی رشته ها ۸۰ درصد شباهت دارند و این شبیه هم را اذغام می کند.
ب) در صورتی که لزوماً بخواهیم همه ی گونه ها را حفظ کنیم و درخت نهایی شامل تقسیم هایی باشد که دقیقاً در همه آمده باشد، یعنی اشتراک مجموعه ها، آن گاه می توان درخت اجماعی اکید را ساخت. به این صورت که یکی از درخت ها را در نظر می گیریم و هر split ای که در بقیه نبود را حذف می کنیم. نحوه حذف هم با منقبض کردن یال متناظرش است. این کار را تا زمانی انجام می دهیم که هیچ split غیر اشتراکی باقی نماند. درخت به دست آمده حتماً اشتراکی می شود.

۲ پاسخ سوال: هم تراز ۲

الف) امتیاز gap در این سوال برابر ۳- است. این امتیاز به صورت خطی حساب شده است و با امتحان کردن اعداد جدول به درستی محاسبه شدند.

ب) هم تراز ۲ محلی، زیرا اعداد سطر و ستون اول صفر هستند و هزینه ای برای آن ها پرداخت نمی کنیم.
ج)



	-	A	T	C	G
-		-۳	-۳	-۳	-۳
A	-۳	۰	۴	۰	-
T	-۳	۴	۰	-	۰
C	-۳	۰	-	۰	۵
G	-۳	-	۰	۵	۰

۳ پاسخ سوال ۳: ماتریس فاصله ها

(الف)

برای متریک بودن شروط زیر را بررسی می کنیم:

$$M_{ij} = M_{ji}$$

$$M_{ii} = ۰$$

$$M_{ij} + M_{jk} \geq M_{ik}$$

با بررسی این شروط برای ماتریس داده شده، متوجه می شویم ماتریس متریک است.
(ب) برای ابرمتریک بودن باید شرط سه نقطه برقرار باشد

$$M_{ik} = M_{jk} \geq M_{ij}$$

با بررسی شرط فوق برای هر سه تایی می بینیم که شرط سه نقطه برقرار است.

$$\text{for } a, b, c : bc = ac \geq ab$$

$$\text{for } a, b, d : ad = bd \geq ab$$

$$\text{for } b, c, d : bc = bd \geq cd$$

(ج) برای جمعی بودن باید شرط چهار نقطه برقرار باشد.

$$M_{ik} + M_{jl} = M_{il} + M_{jk} \geq M_{ij} + M_{kl}$$

با بررسی ماتریس داده شده، می بینیم که جمعی است.

$$bc + ad = bd + ac \geq ab + cd$$

$$۸ + ۸ = ۸ + ۸ \geq ۲ + ۶$$

(د) درخت UPGMA:

۱. کمترین دو فاصله را در نظر بگیر.

۲. فاصله را تقسیم بر ۲ کن و درخت را بساز.

۳. نود جدید را بساز و فاصله ها را مشخص کن.

۴. آپدیت ماتریس، فاصله های جدید: حاصل جمع تقسیم بر ۲



	a	b	c	d
a	۰	۲	۸	۸
b	۲	۰	۸	۸
c	۸	۸	۰	۶
d	۸	۸	۶	۰

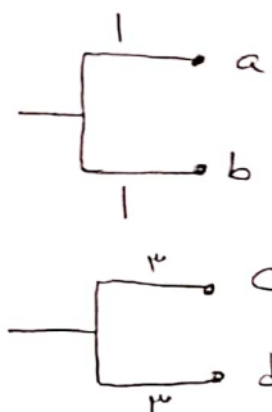
فاصله ab با اندازه ۲ کمترین فاصله را دارد.



آپدیت ماتریس:

	(a,b)	c	d
(a,b)	۰	۸	۸
c	۸	۰	۶
d	۸	۶	۰

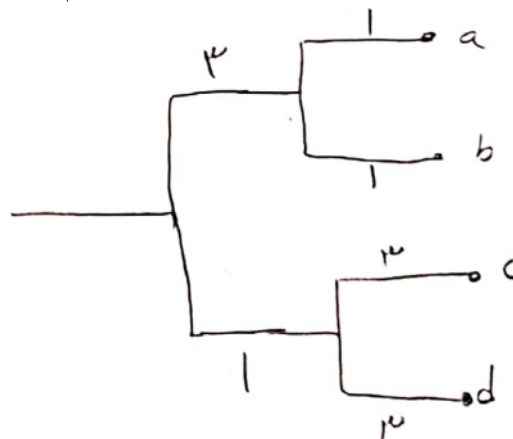
فاصله cd با اندازه ۶ کوتاه‌ترین است. بنابراین داریم:



آپدیت ماتریس:

	(a,b)	(c,d)
(a,b)	۰	۸
(c,d)	۸	۰

حال دو کلاستر کوتاه‌ترین فاصله را دارند و به هم متصل می‌کنیم.



ه) با توجه به اینکه ماتریس ابرمتریک Additive هم است، می‌توان از هر دو الگوریتم UPGMA و Neighbour-Joining استفاده کرد. اما چون UPGMA از $O(n^2)$ و NJ از $O(n^3)$ است، بهتر است از UPGMA استفاده شود.

۴ سوال ۴: هم تراز ی چند گانه

الف) روش Center-Star

ابتدا بین هر جفت رشته هم تراز ی انجام می دهد. اگر m رشته داشته باشد $\binom{m}{2}$ تا هم تراز ی دو به دو انجام می دهد. سپس رشته ای که کم ترین فاصله با بقیه را دارد به عنوان مرکز انتخاب می کند و بقیه را با آن هم تراز می کند و در صورت نیاز به قبلی ها gap اضافه می کند که از $O(m)$ عمل است.

ب) روش ClustalW

در این روش نیز ابتدا یک ماتریس به ازای هر جفت رشته پر می شود که فاصله $1 - \frac{y}{x}$ در آن نوشته شده است که $\frac{y}{x}$ نسبت ستون های مساوری به ستون های غیر فاصله است. بنابراین $\binom{m}{2}$ تا هم تراز ی داریم و ماتریس $m \times m$ پر می شود. سپس الگوریتم NJ درخت متناظر این ماتریس را می سازد. سپس هم تراز ی پرو فایل ها بر اساس درخت انجام می شود که از $O(m)$ عمل است.

۵ سوال ۵: درست نمایی بیشینه

$$\prod_{i=1}^n [\frac{1}{2} L_i(R, \circ) + \frac{1}{2} L_i(R, \vee)]$$

برای ویژگی ۱:

$$L_1(R, \circ) = L_1(T_1, \circ) Pr[T_{1,1} = \circ | R = \circ] + L_1(T_1, \vee) Pr[T_{1,1} = \vee | R = \circ] + L_1(I_1, \circ) Pr[I_{1,1} = \circ | R = \circ] + L_1(I_1, \vee) Pr[I_{1,1} = \vee | R = \circ]$$

$$L_1(I_1, \circ) = L_1(T_2, \vee) Pr[T_{2,1} = \vee | I_1 = \circ] + L_1(T_2, \circ) Pr[T_{2,1} = \circ | I_1 = \circ]$$

$$L_1(I_1, \vee) = L_1(T_2, \vee) Pr[T_{2,1} = \vee | I_1 = \vee] + L_1(T_2, \circ) Pr[T_{2,1} = \circ | I_1 = \vee]$$

احتمال یال ها:

$$P(R, T_1) = p^2(1-p)^2 = \frac{2}{9} = \frac{1}{3}$$

$$P(R, I_1) = p^2(1-p)^2 = \frac{2}{9} = \frac{1}{3}$$

$$P(I_1, T_2) = p^2(1-p)^2 = \frac{2}{9} = \frac{1}{3}$$

$$P(I_1, T_2) = p^2(1-p)^2 = \frac{2}{9} = \frac{1}{3}$$

بنابراین:

$$L_1(R, \circ) = 1 * \frac{1}{3} + [1 * \frac{1}{3} + (1 - \frac{1}{3})](1 - \frac{1}{3}) + [(1 - \frac{1}{3}) + \frac{1}{3}] * \frac{1}{3} = \frac{4}{3}$$

$$L_1(R, \vee) = 0$$

بنابراین حاصل ضرب گفته شده در ابتدای این سوال ، برای ویژگی اول به صورت زیر است:

$$\frac{1}{2} * \frac{4}{3} = \frac{2}{3}$$