Xiheng He

Lisanne Friedrich

Exercises for Algorithmic Bioinformatics II

Assignment 7

Xiheng He

December 2021

Exercise 1 (PAM Matrix, 10P):

Calculate the $PAM_{i,j}$ values of a 2-PAM matrix from the given count matrix:

$f_{i,j}$	A	B	C	D
A	0	3	7	9
B	3	0	5	1
C	7	5	0	8
D	9	1	8	0

$$f_A = \sum_{b \neq A} f_A = 3 + 7 + 9 = 19$$

$$f_B = \sum_{b \neq B} f_B = 3 + 5 + 1 = 9$$

$$f_C = \sum_{b \neq C} f_C = 7 + 5 + 8 = 20$$

$$f_D = \sum_{b \neq D} f_D = 9 + 1 + 8 = 18$$

$$f = \sum_{a} f_a = 19 + 9 + 20 + 18 = 66$$

$\because p_A, p_B, p_C, p_D$ are not given, assume $p_A = p_B = p_C = p_D = 0.25$

$$\because M_{ab} = \frac{1}{100f} \cdot \frac{f_{ab}}{pa} \wedge M^2_{ab} = \sum_c M_{ac} \cdot M_{cb} = M^2_{ab} \wedge PAM(2)_{i,j} = \log \frac{M^2(i,j)}{f(j)}$$

M_{ab}	A	B	C	D
A	$\frac{1631}{1650}$	$\frac{1}{550}$	$\frac{7}{1650}$	$\frac{3}{550}$
B	$\frac{1}{550}$	$\frac{547}{550}$	$\frac{1}{330}$	$\frac{1}{1650}$
C	$\frac{7}{1650}$	$\frac{1}{330}$	$\frac{163}{165}$	$\frac{4}{825}$
D	$\frac{3}{550}$	$\frac{1}{1650}$	$\frac{4}{825}$	$\frac{272}{275}$

$\Longrightarrow$

M^2_{ab}	A	B	C	D
A	0.9762	0.0036	0.0084	0.0108
B	0.0036	0.9890	0.0060	0.0012
C	0.0084	0.0060	0.9760	0.0096
D	0.0108	0.0012	0.0096	0.9784

$\Longrightarrow$

$PAM(2)_{i,j}$	A	B	C	D
A	-1.289	-3.398	-3.377	-3.222
B	-3.722	-0.959	-3.523	-4.176
C	-3.354	-3.176	-1.312	-3.273
D	-3.245	-3.875	-3.319	-1.265