Assignment 3.

1. (a). ①. Choose one gap position in 4 position. which is ${}_4C_1$

②. each orther 3 position has 4 situation (A, C, G or T).

$${}_4C_1 \times 4^3 = 4 \times 4^3 = 256.$$

(b).

S_1 | S_2.

1-gapped 3-mers	No.of occurrence		1-gapped 3-mers	No.of occurrence
*CCG	1		*ACG	1
*CGA	1		*CGA	1
*CGT	1		*CGC	1
*GAC	1		*GAC	2
*GTC	1		*GCG	1
*TCC	1		A*GC	1
C*AC	1		AC*C	1
C*GA	1		ACG*	1
C*TC	1		C*AC	2
CC*A	1		C*CG	1
CCG*	1		CG*C	2
CG*C	2		CG*G	1
CGA*	1		CGA*	2
CGT*	1		CGC*	1
G*CC	1		G*CG	1
G*GT	1		G*GA	1
GC*T	1		GA*G	1
GCG*	1		GAC*	1
GT*C	1		GC*A	1
GTC*	1		GCG*	1
T*CG	1			
TC*G	1			
TCC*	1			

(c). Similarity = 1 (due to *CGA) + 2 (*GAC) + 2 (C*AC) + 4 (CG*C)

+ 2 (CGA*) + 1 (GCG*)

= 12

(d).

s1\s2	ACGC	CGAC	CGAC	CGCG	GACG	GCGA
CCGA	2	3	3	3	4	1
CGAC	3	0	0	2	4	4
CGTC	3	1	1	2	4	4
GCGT	2	4	4	4	3	1
GTCC	3	3	3	3	2	3
TCCG	3	4	4	2	2	3

(e). With the result in (d),

We know • if there is one gap, thus if there is more than one mismatch, the score = 0.

- If there is 1 mismatch, score = 1 because there is only one situation to led the gap to be the position of mismatch.
- If there is 0 mismatch, score = 4, because there is 4 situation that we place the gap.

We can now build the similarity score table below.

s1\s2	ACGC	CGAC	CGAC	CGCG	GACG	GCGA
CCGA	0	0	0	0	0	1
CGAC	0	4	4	0	0	0
CGTC	0	1	1	0	0	0
GCGT	0	0	0	0	0	1
GTCC	0	0	0	0	0	0
TCCG	0	0	0	0	0	0

Therefore, the score = $4+1+4+1+1+1 = 12$

(f). ① When calculate the mismatch of $(k+g)$-mers of two sequence, this can be speeded up by using XOR operations.

② Avoid considering all $(k+g)$-mer, using a tree to represent the all $(k+g)$-mer, and if there is more than g mismatch, stop going further.

2. (a). $Pr(X_1=1) = Pr(X_1=1, X_2=0) + Pr(X_1=1, X_2=1)$ — Law of total probability.

$Pr(Y=1 | X_1=1, X_2=1) \cdot Pr(X_1=1, X_2=1) = Pr(X_1=1, X_2=1, Y=1)$

Thus, $Pr(Y=0 | X_1=1, X_2=1) = 1 - Pr(Y=1 | X_1=1, X_2=1)$

$$= 1 - \frac{Pr(X_1=1, X_2=1, Y=1)}{Pr(X_1=1, X_2=1)}$$

$$= 1 - \frac{Pr(X_1=1, X_2=1, Y=1)}{Pr(X_1=1) - Pr(X_1=1, X_2=0)}$$

(b). 卧槽.

(c). $Pr(X_1|Y) \cdot Pr(X_2|Y) = Pr(X_1, X_2 | Y)$.

(d). Sample: $Pr_1 = Pr(X_1=0, X_2=0 | Y=0) = 0.6$

$Pr_2 = Pr(X_1=0, X_2=1 | Y=0) = 0.2$

$Pr_3 = Pr(X_1=1, X_2=0 | Y=0) = 0.2$

$Pr_4 = Pr(X_1=1, X_2=1 | Y=0) = 0$

$Pr_5 = Pr(X_1=0, X_2=0 | Y=1) = 0$.

$Pr_6 = Pr(X_1=0, X_2=1 | Y=1) = 0$

$Pr_7 = Pr(X_1=1, X_2=0 | Y=1) = 0$

$Pr_8 = Pr(X_1=1, X_2=1 | Y=1) = 1$

Prove:

$Pr(X_1=1 | Y=0) = Pr_3 + Pr_4 = 0.2$.

$Pr(X_2=1 | Y=0) = Pr_2 + Pr_4 = 0.2$.

While $Pr(X_1=1|Y=0) \cdot Pr(X_2=1|Y=0) = 0.04 \neq Pr(X_1=1, X_2=1 | Y=0)$.

Thus we know X_1 and X_2 are not conditinally independent.

(e) No.

Gives conditional independent X_1, X_2 that $(X_1|Y) \cdot (X_2|Y) = (X_1 \text{ and } X_2 | Y)$.

Let's set $Pr(X_1=1 | Y=0) = 0.2$

$Pr(X_2=1 | Y=0) = 0.2$.

$Pr(X_1=1 | Y=1) = 1$.

$Pr(X_2=1 | Y=1) = 1$. and $Pr(Y=1) = 0.9$.

And $\Pr(X_1=1) = \Pr(X_1=1 \text{ and } Y=0) + \Pr(X_1=1 \text{ and } Y=1)$

$= 0.2 \times 0.1 + 1 \times 0.9 = 0.92.$

$\Pr(X_2=1) = \Pr(X_2=1 \text{ and } Y=0) + \Pr(X_2=1 \text{ and } Y=1)$

$= 0.2 \times 0.1 + 1 \times 0.9 = 0.92.$

$\Pr(X_1=1 \text{ and } X_2=1) = \Pr(X_1=1, X_2=1 \mid Y=0) \cdot \Pr(Y=0)$

$+ \Pr(X_1=1, X_2=1 \mid Y=1) \cdot \Pr(Y=1)$

$= 0.2 \times 0.2 \times (1-0.9) + 1 \times 1 \times 0.9$

$= 0.9004$

And $\Pr(X_1=1) \cdot \Pr(X_2=2) \neq \Pr(X_1=1, X_2=2).$

Thus, X_1, X_2 is conditional independent on Y but not unconditionally independent.

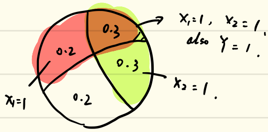
(f). No. Gives $\Pr(X_1) \cdot \Pr(X_2) = \Pr(X_1, X_2).$

Set $\Pr(X_1=1) = 0.5$

$\Pr(X_2=1) = 0.6$

$\Pr(Y=1) = 0.3$

And $\Pr(X_1=1, X_2=1 \mid Y=1) = 1.$



In this situation, $\Pr(X_1=1, Y=0)$

$= \Pr(X=1) - \Pr(X_1=1, Y=1)$

$= \Pr(X=1) - \left(\Pr(X_1=1, X_2=1, Y=1) + \Pr(X_1=1, X_2=0, Y=1)\right)$

$= \Pr(X=1) - \left(\Pr(X_1=1, X_2=1 \mid Y=1) \cdot \Pr(Y=1) + \Pr(X_1=1, X_2=0 \mid Y=1) \cdot \Pr(Y=1)\right)$

$= 0.5 - (1 \times 0.3 + (1-1) \cdot 0.3) = 0.2$

Similarly $\Pr(X_2=1, Y=0) = 0.3$

$\Pr(X_1=1 \mid Y=0) = \dfrac{\Pr(X_1=1, Y=0)}{\Pr(Y=0)} = \dfrac{2}{7}$, $\Pr(X_2=1 \mid Y=0) = \dfrac{3}{7}$

And $\Pr(X_1=1, X_2=1 \mid Y=0) = \Pr(X_1=1, X_2=1, Y=0) / \Pr(Y=0)$

$= [\Pr(X_1=1, X_2=1) - \Pr(X_1=1, X_2=1, Y=1)] / 0.7$

$= [0.5 \times 0.6 - \Pr(X_1=1, X_2=1 \mid Y=1) \cdot \Pr(Y=1)] / 0.7$

$= (0.3 - 0.3)/0.7 = 0 \neq \Pr(X_1=1 \mid Y=0) \cdot \Pr(X_2=1 \mid Y=0)$

Thus X_1 and X_2 are not conditional independent.

CSCI3220 2018-19 First Term Assignment 3

I declare that the assignment here submitted is original except for source material explicitly acknowledged, and that the same or closely related material has not been previously submitted for another course. I also acknowledge that I am aware of University policy and regulations on honesty in academic work, and of the disciplinary guidelines and procedures applicable to breaches of such policy and regulations, as contained in the following websites.

University Guideline on Academic Honesty: http://www.cuhk.edu.hk/policy/academichonesty/

Student Name: ZHANG Chongzhi
Student ID : 1155077072