

國立中央大學98學年度碩士班考試入學試題卷

所別：系統生物與生物資訊研究所碩士班 一般生 科目：計概與資料結構 共 1 頁 第 1 頁  
\*請在試卷答案卷(卡)內作答

人體有 3 萬個基因( $G_1, G_2, \dots, G_n$ )，使用微陣列技術取得  $t_1, t_2, t_3$  三個不同時間這 3 萬個基因的基因表現資料(每個基因的表現資料為一實數)。

試以回答以下問題：

1. 請說明最簡單儲存這些基因表現的資料結構名稱與範例。(10 points)

2. 任兩個基因  $X, Y$  可以因為 3 個不同時間的表現量定義一個距離，例如

$$D(X, Y) = \sqrt{(X_{t_1} - Y_{t_1})^2 + (X_{t_2} - Y_{t_2})^2 + (X_{t_3} - Y_{t_3})^2}$$

請寫出一個有效的演算法名稱與虛擬碼在一台擁有 500Mb RAM 的機器上(不使用虛擬記憶體)排序這些距離。(30points)

3. 且寫出一個有效的虛擬碼能在  $O(n)$  時間算出這些距離的平均值。(20points)

4. 將小於平均值的距離刪除，大於平均值的基因配對連接起來，將形成一個網路。提出一個有效的的資料結構名稱與範例可儲存這個網路。(10 points)

5. 給定其中某個基因  $A$ ，提出兩種有效的演算法名稱與虛擬碼可以找出與  $A$  相關的所有基因。[相關：(有直接/間接連接)/經過某個路徑能連接到] (30 points)

參考用