

BIOINFORMATICA
Compito scritto del 7 Luglio 2008

1) Si consideri il seguente problema computazionale:

INPUT: un vettore A di n numeri.

OUTPUT: numero delle posizioni del vettore in cui sono presenti numeri che sono ripetuti almeno due volte in A .

1. Si descriva a parole, il piu' possibile rigorosamente ma sinteticamente, un algoritmo efficiente che risolva il problema posto. Si motivi la scelta di tale algoritmo.
2. Se ne scriva lo pseudocodice.
3. Se ne analizzi la complessita' **motivando la risposta**.
4. Si applichi l'algoritmo proposto ai vettori $A1 = 1\ 2\ 4\ 2\ 5$ e $A2 = 3\ 3\ 3\ 1\ 6$ mostrando i passi eseguiti. Si verifichi che esso restituisca correttamente il valore 2 per $A1$ e il valore 3 per $A2$.

2) Si mostri la matrice di programmazione dinamica per il calcolo dell'allineamento semiglobale (anche detto allineamento "best fit") tra la sequenza $X = AGCTGCTATGA$ e la sequenza $Y = TGTT$. Si individui sulla matrice suddetta il miglior allineamento di Y con una sottostringa di X , e lo si mostri.

Si consideri score -2 per gli indel, -1 per i mismatch, e 1 per i match.