

Esame 22/06/2017

Andrea Passerini
passerini@disi.unitn.it

Programma python

Scrivere un programma python che:

- prenda in ingresso un file contenente i livelli di espressione di 5147 geni per pazienti affetti da ALL o AML (ogni riga è un paziente), ed un numero n di geni da considerare
- stampi gli n geni per cui la differenza tra espressione media in pazienti ALL ed espressione media in pazienti AML è massima

Esempio file

```
bioC      ADAR      ...  GYPA_3  Z78285_f_at  label
88         15091     ...  191     -37          ALL
283        11038     ...  76      -14          ALL
09         16692     ...  228     -41          ALL
...
9          39538     ...  379     -60          AML
```

Esempio esecuzione

```
> python discriminative_genes.py
Inserire nome file: leukemia.txt
Inserire numero geni: 10
CST3 : 8934.932
FTL : 8710.694
MPO : 8162.000
AZU1 : 7653.641
IL8_2 : 7024.603
GPX1 : 6037.786
S100A9 : 5833.323
VIM : 5814.156
DF : 5263.003
TCL1A : 4925.465
```

Programma python: suggerimento

Si possono implementare 5 funzioni separate:

1. Una che legga il file e restituisca un dizionario che mappa da classe (ALL o AML) a dizionario di geni, con per ogni gene la lista dei valori di espressione dei pazienti della classe
2. Una che data una lista di valori restituisca la loro media
3. Una che dato il dizionario di dizionari letto, produca un lista di coppie gene, differenza tra le medie dei suoi valori nei due tipi di pazienti (usando la funzione 2)

4. Una che data la lista di differenze e il numero n di geni da scegliere, ordini la lista e stampi gli n geni con differenza massima
5. una che realizzi il programma richiesto usando le funzioni di cui sopra

Shell: esercizio #1

Calcolare quante sequenze nel file `sequences.fasta` contengono:

- Una metionina (M) ad inizio sequenza, seguita a qualunque distanza da un pattern formato da una coppia acido aspartico (D) e glicina (G) oppure una coppia alanina (A) seguita da una glicina (G) o una istidina (H), ripetuto due o tre volte e seguito da una glicina (G)
- Ad esempio **MSJRKLDGAHGTJWRTREY**

Risultato atteso

9

Shell: esercizio #2

Dato il file `leukemia.txt`, estrarre:

- il valore di espressione massimo tra tutti i geni in tutti i pazienti
- il valore di espressione minimo tra tutti i geni in tutti i pazienti

Risultato atteso

massimo: 71369 minimo: -10162

Modalita' di esecuzione e consegna

1. Avviare la macchina in modalita' ESAME
2. Autenticarsi con nome utente `sci-esame` e password fornita dal docente
3. Il testo del compito ed i file necessari si trovano in una cartella `Testo` sul Desktop
4. Realizzare il programma python come file `programma.py` e scrivere gli esercizi da linea di comando in un file di testo `linea_di_comando.txt`
5. Creare sul Desktop una cartella con `nome_cognome` e metterci i due file realizzati.
6. Eseguire il logout ma NON spegnere la macchina