Report: Assignment-04

Vedpal Jangir (19299)

MTech-AI

1. **Code summary: -**

♦ At the beginning, all the data is initially extracted from files.

♦ Use **delta** = 5

♦ Class **Rank** has methods **lcr, fci, fcfi, fcli** return last column rank, first col Index, first col first Index, first col last Index, respectively.

♦ The **exact_match** function is for exact match and **mis_match** function is for checking mismatches and allowing up to 2 mismatched characters string.

♦ For the **probability function**, we use the Bernoulli distribution as each read that got mapped to the two genes (R and G) for the given configurations. Take **p** as the probability of R(success) then

$$P = C^{r+g}_{r} \ (p)^r \ (1-p)^g$$

Now, as the probability is much less, we calculate the log-likelihood of each configuration, as it gives the same result.

2. **Result: -**

   2.1 Exact Match -

R	135	61	67	128	275	358
G	135	185	75	123	328	358

   2.2 allowed mismatch -

R	181	89	94	178	332	444
G	181	239	148	159	398	444

   2.3 Time: -

File loading time = 3.5 to 4 minutes approximate.

Runtime = 1.5 to 2hr approximately

2.4 Memory: - 8 GB with swap memory of 2 GB**.**

2.5 Probability: - log likelihood

P1	P2	P3	P4
-67.227	–inf	-27.7429	-86.6541

Configuration with 33%,33%,100%,100% the maximum likelihood among the four Configurations**.**