

Algorithms in Genome Research
Winter 2017/2018

Exercises

Number 10, Discussion: 2018 January 19

Exercises by M. Sc. Linda Sundermann.

1. The lineage relation matrix Z

- Does the following matrix Z encode a proper phylogenetic tree? If not, which entries prevent the matrix from encoding a tree?

-1	1	1	1
-1	-1	1	-1
-1	-1	-1	1
-1	-1	-1	-1

- The following matrix Z does not encode a proper tree. To get a proper tree, it's sufficient to change one single entry. Which one is it and how do you have to change it?

-1	1	1	1	1	1
-1	-1	0	1	1	0
-1	-1	-1	1	1	0
-1	-1	-1	-1	1	0
-1	-1	-1	-1	-1	0
-1	-1	-1	-1	-1	-1

- When checking whether a matrix Z encodes a proper tree, all triplets $(Z_{k,k'}, Z_{k,k''}, Z_{k',k''})$, with $1 \leq k < k' < k'' < K$, need to be considered. How many different triplets do exist?

2. Lineage-based subclonal reconstruction

A tumor samples was sequenced and investigated. You know that there are

- three lineages with the frequency 100 %, 80 % and 30 %,
- two segments with the following allele-specific copy numbers: $c_A = (1, 1.3)$, $c_B = (1, 1)$,
- four SNVs, the first two lie on segment 0, the last two on segment 1,
- the following read counts for the SNVs: $V = (28, 11, 39, 28)$ and $D = (70, 70, 81, 81)$
- Find the subclonal reconstruction that maximizes the log likelihood.

frequency vector ϕ

lineage relation matrix Z

matrices ΔC_A and ΔC_B for appearing copy number changes

matrices ΔS , ΔS_A and ΔS_B for appearing SNVs

3. Variables and constraints of SNVs

- Define the variables ΔS , ΔS_A and ΔS_B for the MILP¹?
- Write down the constraint for the infinite site assumption for an SNV j .

¹mixed integer linear programming