

### 3. Exercise “Bioinformatische Methoden in der Genomforschung”

Sebastian Böcker, Martin Hoffmann

Assigned: 08.11.2021

Due: 15.11.2021

#### Exercise 1 (2 Points)

What are the tasks of a microarray data analysis? Which steps does it usually consist of?

#### Exercise 2 (2 Points)

Significance (p-values) is often calculated when statistically analysing expression data. Complete the following sentence: A p-value of 0.01 when testing for differential gene expression means ...”.

#### Exercise 3 (6 Points)

Given the following matrix of gene expression values measured for four genes on four consecutive days:

	Day 1	Day 2	Day 3	Day 4
Gene 1	0.564	-0.038	-0.561	-1.315
Gene 2	0.606	0.621	-0.83	-1.681
Gene 3	-0.555	-0.224	0.673	0.78
Gene 4	0.238	-0.764	-1.371	-1.868

1. Calculate the distance matrices for the following distances:

(a) Euclidian

$$d_E(x_i, x_j) = \sqrt{\sum_k (x_{ik} - x_{jk})^2}$$

(b) Manhattan

$$d_M(x_i, x_j) = \sum_k |x_{ik} - x_{jk}|$$

(c) Correlation

$$d_C(x_i, x_j) = 1 - \frac{\sum_k ((x_{ik} - \mu_i)(x_{jk} - \mu_j))}{n\sigma_i\sigma_j}$$

where  $x_{ik}$  stands for the expression value of the  $i$ -th gene on the  $k$ -th day,  $\mu_i$  for the mean,  $\sigma_i$  for the standard deviation of the expression value of the  $i$ -th gene and  $n$  for the number of days.

$$\sigma_i = \sqrt{\frac{1}{n} \sum_k (x_{ik} - \mu_i)^2}$$

2. Normalise the following input matrix according to this *Standardisation approach*:

**Step 1.** For every value in a row, use transformation  $x \mapsto \frac{x - \mu}{\sigma}$ , where  $\mu$  is the mean and  $\sigma$  is the standard deviation of the row values.

**Step 2.** For every value in a column use the same transformation, except now calculate mean and standard deviation using column value

Repeat Step 1.

Repeat Step 2.

etc ...

How does the matrix look after one,two, three steps? (Bonus: How does the matrix look after 10, 100, 1000 steps?) What can you observe?