

### 3. Übung zur Vorlesung “Bioinformatische Methoden in der Genomforschung”

Sebastian Böcker, Martin Engler

Ausgabe: 12.11.2015

Abgabe: 19.11.2015

#### Aufgabe 1 (2 Punkte)

Ist das Shortest-Common-Supersequence-Problem für die Microarray-Produktion wichtig? Motivieren Sie Ihre Antwort.

#### Aufgabe 2 (2 Punkte)

Häufig werden bei der statistischen Analyse von Expressionsdaten Signifikanzen (p-values) berechnet. Vervollständigen Sie den folgenden Satz: “Ein p-value (Signifikanz) von 0.01 beim Test auf differentielle Genexpression bedeutet...”.

#### Aufgabe 3 (6 Punkte)

Gegeben sei die Matrix der Genexpressionswerte gemessen für vier Gene an vier (aufeinanderfolgenden) Tagen:

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

Normalisieren Sie die Eingabematrix mit dem folgenden *Standardisierungsansatz*:

**Schritt 1.** Für jeden Wert in der Zeile verwenden Sie die Transformation  $x \mapsto \frac{x-\mu}{\sigma}$ , wobei  $\mu$  der Mittelwert und  $\sigma$  die Standardabweichung der Zeilenwerte ist.

**Schritt 2.** Für jeden Wert in der Spalte verwenden Sie dieselbe Transformation mit dem Unterschied, dass der Mittelwert und die Standardabweichung nun über die Spalten berechnet werden.

Wiederholen Sie den Schritt 1.

Wiederholen Sie den Schritt 2.

usw ...

Wie sieht die Matrix nach einem, zwei, drei Schritte aus? (Zusatzaufgabe: Wie sieht die Matrix nach 10, 100, 1000 Schritten aus?) Was können Sie dabei beobachten?

**Aufgabe 4** (5 Punkte)

Gegeben sei die Matrix der Expressionswerte eines Gens von zehn Personen der Kontrollgruppe C und den Testgruppen A und B (nach Einnahme des Medikaments A bzw. B):

A	1.4436	2.0271	0.8835	-0.2830	0.4962	1.8128	0.3567	1.5766	1.0744	1.8620
B	1.0196	1.9012	1.3961	3.4540	2.4126	2.6895	1.4709	1.7657	1.8638	2.5492
C	0.8781	1.1078	0.4171	0.4260	1.0524	1.3611	2.2927	0.6666	1.0937	0.9588

Die Null-Hypothese  $H_0$  des t-Tests besagt, dass die Mittelwerte zwei normal verteilter Stichproben gleich sind:  $\mu_1 = \mu_2$ , die Alternativhypothese  $H_1$  besagt, dass gilt:  $\mu_1 \neq \mu_2$ . Sei  $n$  die Stichprobengröße,  $D$  die Differenz zweier Messreihen und  $t = \sqrt{(n)} \frac{\mu_D}{\sigma_D}$  die Teststatistik. Wir akzeptieren  $H_0$ , falls  $|t| \leq 2.2622$  (Wert der t-Verteilung mit neun Freiheitsgraden und einem Signifikanzniveau von 5%). Welches der beiden Medikamente war ein Placebo?