

Übungen zur Vorlesung Sequenzanalyse I

Universität Bielefeld, WiSe 2009/2010

Prof. Dr. Jens Stoye · Dipl.-Inform. Nils Hoffmann

<http://wiki.techfak.uni-bielefeld.de/gi/GILectures/2009winter/SequenzAnalyse>

Blatt 7 vom 26.11.2009

Abgabe in einer Woche vor Beginn der Vorlesung.

Aufgabe 1 Äquivalente Score-Funktion

(3 Punkte)

1. Erkläre, was es bedeutet, wenn eine Score-Funktion *äquivalent* zu einer Kostenfunktion ist.
2. Gegeben sei die folgende Kostenfunktion zur Berechnung der Edit-Distanz:

$$\text{cost}(\mathcal{C}) = 0$$

$$\text{cost}(\mathcal{S}_{a,c}) = 1$$

$$\text{cost}(\mathcal{I}_c) = \text{cost}(\mathcal{D}_c) = 3$$

Gib eine äquivalente Score-Funktion an!

Aufgabe 2 Scorematrizen

(2 Punkte)

Informiere Dich über die *PAM*(*t*) und *BLOSUM*(*s*) Matrizen.

1. Wie wurden diese Matrizen erstellt?
2. Worin unterscheiden sich die Matrizen?

Aufgabe 3 Substitutions-Scorematrix

(3 Punkte)

Gegeben sei $\Sigma = \{a,b,c\}$. Erstelle aus den folgenden Sequenzen eine Substitutions-Scorematrix:

babca
abcca
bacca
bcbba
aaaac

Bestimme dazu aus diesen Sequenzen die

1. Observationsmatrix M^t
2. Hintergrundverteilung π
3. Likelihood-Ratio $\left(\frac{M^t(x,y)}{\pi_x \pi_y}\right)$,

um schließlich die log-Likelihood Scorematrix $\left(s^t(x,y)\right) := \left(\log_2\left(\frac{M^t(x,y)}{\pi_x \pi_y}\right)\right)$ anzugeben.

Aufgabe 4 Ein Parser für Score-Matrizen

(2 Punkte)

Implementiere eine JAVA-Klasse `ScoreMatrix`, um log-Likelihood Scorematrizen einlesen und verwalten zu können. Das Format einer Score-Matrix ist in diesem Fall das einer quadratischen Matrix. Die erste Zeile der Datei besteht aus den Symbolen des Alphabets, getrennt durch Leerzeichen (z.B. `\t` (tabstop)); gleichzeitig ist dies die Beschriftung der Spalten und Zeilen. Zur Verdeutlichung ein Beispiel:

A	R	N
5.3	-8.2	-5.01
-3.45	10.3	-4.32
-1.04	-5.12	8.32

Implementiere die Methode `public double getScore(String a, String b)`, mit der man den Substitutionsscore zweier Symbole des Alphabets abfragen kann. Diese Methode sollte eine Ausnahme vom Typ `IllegalArgumentException` werfen, wenn eines oder beide Symbole unbekannt sind.