

# Übungen zur Vorlesung Grundlagen der Sequenzanalyse

Universität Bielefeld, WS 2006/07

Dr. Sven Rahmann · Dipl.-Bioinf. Katharina Jahn

<http://gi.cebitec.uni-bielefeld.de/teaching/2006winter/sequenzanalyse/>

**Blatt 1 vom 19.10.2006**

**Abgabe am 26.10.2006 vor der Vorlesung um 8:30 in H3**

**Aufgabe 1** Auf jeder Menge  $X$  lässt sich eine Abbildung  $d : X \times X \rightarrow \mathbb{R}$  definieren durch

$$d(x, y) := \begin{cases} 0, & \text{wenn } x = y, \\ 1, & \text{wenn } x \neq y. \end{cases}$$

Weise nach, dass es sich bei  $d$  um eine Metrik auf  $X$  handelt.

**Aufgabe 2** Wir betrachten die Ebene  $\mathbb{R}^2$  mit Punkten  $x = (x_1, x_2)$  und  $y = (y_1, y_2)$ , sowie die Metriken (ein Nachweis der Metrik-Eigenschaften ist nicht erforderlich; die Dreiecksungleichung ist nicht trivial!)

$$\begin{aligned} d_1(x, y) &:= |x_1 - y_1| + |x_2 - y_2| && \text{(Manhattan-Distanz),} \\ d_2(x, y) &:= \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2} && \text{(euklidische Distanz),} \\ d_\infty(x, y) &:= \max\{|x_1 - y_1|, |x_2 - y_2|\} \end{aligned}$$

Zeichne jeweils die Menge der Punkte, die den Abstand 1 vom Nullpunkt haben.

**Aufgabe 3** Gegeben sind die Sequenzen  $s_1 = \text{ACGATACTAG}$ ,  $s_2 = \text{AGGCATTGA}$ ,  $s_3 = \text{AGGCATTGA}$ ,  $s_4 = \text{CGATACTAGA}$ . Berechne sämtliche paarweisen Hamming-Distanzen zwischen diesen Sequenzen. Schreibe eine Funktion `HammingDistanz` in einer Sprache Deiner Wahl, die diese Aufgabe übernimmt. (Programmieraufgaben dienen der Arbeitserleichterung und müssen nicht abgegeben werden!)

**Aufgabe 4** Gegeben sei ein Alphabet  $\Sigma$  der Größe  $|\Sigma| = \sigma$  und eine Sequenz  $s \in \Sigma^n$ . Wie viele Sequenzen gibt es, die eine Hammingdistanz von  $0, 1, 2, \dots$  zu  $s$  haben? Gib Sie eine allgemeine Formel für die Anzahl der Sequenzen mit Hammingdistanz  $k$  an.

**Wichtige Hinweise:** Bitte unbedingt in eine der Übungslisten mit Terminwunsch eintragen und im ekvv für die Übungen und für die Vorlesung registrieren (verschiedene Veranstaltungsnummern!).

Die Klausur kann am Ende mitschreiben, wer bis dahin A&D bestanden, mindestens 60% der Übungspunkte erreicht und mindestens zweimal in den Übungen eine Aufgabe vorgerechnet hat. Zusammenarbeit in Arbeitsgruppen beim Lösen der Aufgaben ist völlig in Ordnung. Abschreiben hingegen widerspricht allen ethischen Grundsätzen und kann zur Exmatrikulation führen. Wir verstehen da keinen Spaß!