



Prof. Hanspeter Herzel
Prof. Nils Blüthgen
Dr. Manuela Benary
Institute for Theoretical Biology

www.sys-bio.net/teaching
h.herzel@biologie.hu-berlin.de
nils.bluthgen@charite.de
manuela.benary@biologie.hu-berlin.de

BIOINFORMATIK SS 2017

Gib deine Lösung bitte am 29.05. in der Vorlesung ab. Alternativ kannst du die Lösung auch bis zum Vorlesungszeitpunkt per E-mail an manuela.benary@charite.de schicken.

1. Falsch Positive

An einem Flughafen wird ein Massenspektrometer aufgestellt, der verbotene Substanzen detektieren soll. Die Sicherheitsbehörden vermuten, dass ca. 0.1% der Passagiere verbotene Substanzen mitführen. Der Detektor erkennt diese mit 98% der Fälle. Wenn keine verbotenen Substanzen dabei sind, so detektiert der Spektrometer das zu 99% richtig.

- (a): Wo groß ist die Falsch-Positiv-Rate und Falsch-Negativ-Rate des Detektors?
- (b): Den Flughafen passieren im Jahr 1 Mio Passagiere. Wie oft schlägt der Detektor Alarm?
- (c): Wie oft detektiert der Spektrometer im Jahr fälschlicherweise einen Passagier, der nichts Verbotenes mitführt?
- (d): Wenn der Detektor anschlägt, wie hoch ist die Wahrscheinlichkeit, dass der Passagier etwas Verbotenes mitführt?

2. Clustern

- (a): Stelle die Distanzmatrizen für die Unterschiede der verschiedenen Samples auf, einmal mit der Manhattan-Distance, einmal mit der Korrelationsdistanz.
- (b): Benutze die Distanzmatrizen zum Clustern der Samples. Benutze dabei Average Linkage. Zeichne jeweils das Dendogram.
- (c): Warum sind die Cluster unterschiedlich? Interpretiere kurz die Unterschiede.

Daten:

sample 1	sample 2	sample 3
1	2	4
2	4	8
3	3	12