

Tutorium "Strategien zur Analyse hochdimensionaler molekularer Daten"

Dozent

PD Dr. Harald Binder, Institut für Medizinische Biometrie und Medizinische Informatik, Universitätsklinikum Freiburg

Dauer:

10 bis 18 Uhr (eine Stunde Pause)

Teilnehmerzahl: maximal 20

Voraussetzungen:

Kenntnis der Grundprinzipien des statistischen Testens und der Regressionsmodellierung. Basiskenntnisse im Umgang mit der Statistikumgebung R. Für die praktischen Übungen bitten wir die Teilnehmer, ihren eigenen Laptop zum Kurs mitzubringen.

Beschreibung:

Neuere molekulare Messtechniken erlauben eine Untersuchung des Zusammenhangs von vielfältigen Vorgängen auf Zellebene mit klinischen Endpunkten bzw. Erkrankungsphänotypen. So lassen sich z.B. bei jedem untersuchten Individuum punktuellen Veränderungen der Erbinformation bestimmen, mit Genexpressions-Microarrays die Transkription der genetischen Information messen, oder mit Techniken zur Quantifizierung der Methylierung epigenetische Mechanismen untersuchen. Aus statistischer Perspektive ist all diesen Techniken gemein, dass sich für jedes Individuum zehntausende oder sogar Millionen von Messwerten ergeben, die zusätzlich zu etablierten klinischen Faktoren in die Modellierung eines Endpunktes oder Phänotyps einbezogen werden sollen. Dieses Tutorium wird, aufbauend auf klassischen biostatistischen Techniken, Strategien zur Analyse derartiger Datenmengen vermitteln. Eine Komponente sind dabei Test-basierte Ansätze, die es erlauben, Listen potentiell interessanter molekularer Größen zu erstellen. Der Fokus liegt hier auf der Kontrolle des Anteils der falsch positiven Ergebnisse. Für die Entwicklung prognostischer oder prädiktiver Signaturen werden im Kontrast dazu die Vorteile multivariabler Regressionsmodelle in Kombination mit regularisierter Schätzung aufgezeigt. Spezifisch wird dabei der Umgang mit komponentenweisen Boosting-Techniken vermittelt, die auch komplexe Ereigniszeit-Endpunkte berücksichtigen können. Die Evaluierung der resultierenden Modelle ist eine wesentliche Komponente jeglicher Analysestrategie. Das Tutorium wird dazu Ansätze zur unverzerrten Schätzung der Vorhersageleistung einführen und gängige Fehler bei der Bewertung illustrieren. Spezifisch für Ereigniszeit-Endpunkte werden Techniken zur Bestimmung von Vorhersagefehlerkurven vermittelt. Ein besonderes Anliegen dieses Tutoriums ist es dabei, aufzuzeigen, wie aus der Kombination all dieser Techniken, d.h. aus der Kombination von Test-basierten Ansätzen, regularisierten Regressionstechniken und von Verfahren zur Evaluation, Analysestrategien für vielfältige Anwendungen mit hochdimensionalen molekularen Daten entwickelt werden können. Die Teilnehmerinnen und Teilnehmer sollen so befähigt werden eigene konkrete Fragestellungen anzugehen.