

## Interdisziplinäres Qualifikationsprogramm für Nachwuchswissenschaftler:innen der Wirtschafts- und Verhaltenswissenschaftlichen Fakultät der Universität Freiburg

### Online-Workshop zum Thema "Verfahren der Künstlichen Intelligenz unter Verwendung von R" (Procedures in "Artificial Intelligence" using R)

#### Dozent, Zeit und Ort

Dozent: PD Dr. Joseph Krummenacher

Zeit: montags 16-18 Uhr; am 18.01.2021; 25.01.21; 01.02.21; 08.02.21; 15.02.21; 22.02.21.

Ort: online; via Zoom

Hinweis: Als Veranstaltungssprache ist Deutsch vorgesehen, auf Wunsch kann der Kurs auf Englisch gehalten werden. (Note that event language will be German, however, the course may be delivered in English if desired.)

#### Anmeldung

Zulassung: vorzugsweise für Postdoktorand/innen und Doktorand/innen der WVF

Anmeldung: per E-Mail bis 7. Januar an [yziegler@psychologie.uni-freiburg.de](mailto:yziegler@psychologie.uni-freiburg.de)

Auswahl: erfolgt nach dem Prinzip first come, first serve

Teilnehmerzahl: 4 bis 12

Voraussetzung: Grundkenntnisse in Regressionsanalyse sowie der Programmierumgebung *RStudio* sind wünschenswert, aber keine Bedingungen. *R* und *RStudio* sind für die akademische Verwendung kostenfrei, für die gängigen Betriebssysteme verfügbar und können unter den URLs <https://cran.r-project.org/> und <https://rstudio.com/products/rstudio/download/> gefunden werden. (*RStudio* baut auf *R* auf und erfordert daher eine vorangehende Installation von *R*).

#### Inhalte des Workshops

Der Begriff der „künstlichen Intelligenz“ hat eine große Verbreitung und in einem Teil der Medien einen nahezu mythischen Charakter erlangt. Alternative Bezeichnungen wie „Machine Learning“, „Statistical Learning“ oder „Statistical Modeling“ verweisen auf die zugrundeliegenden Konzepte und Vorgehensweisen etwa in der mathematischen Optimierung und Modellierung oder der Stochastik und Statistik.

In diesem Kurs werden in der künstlichen Intelligenz eingesetzte Ansätze theoretisch – durch die Diskussion von Konzepten auf der Basis von Regressionsverfahren – und praktisch – durch die Anwendung von Verfahren zur Lösung von Problemen unter Verwendung des Statistikpakets *R* – exploriert. Da dieses Vorgehen einen engen Bezug zur Statistik hat, wird im Kurs häufig auch der Begriff des „statistischen Lernens“ verwendet.

(Detaillierte Angaben zu den geplanten Inhalten finden Sie weiter unten unter *Ablauf des Workshops*.)

WVF-Nachwuchsförderungsprogramm 2020-2021

Koordination: Dr. Yvonne Ziegler

*Literatur* (zur Ergänzung und Vertiefung): James, G., Witten, D., Hastie, T. & Tibshirani, R. (2013). *An Introduction to Statistical Learning with Applications in R*. New York: Springer.

*Folien und Material*: Folien und zusätzliche Materialien (Datendateien, R-Skripte, PDFs) werden auf Ilias verfügbar gemacht.

*Fakultative Übungsblätter*: Zu jedem Thema gibt es ein Übungsblatt mit Problemen, die (auf fakultativer Basis) mithilfe von R gelöst werden können. Musterlösungen werden zur Verfügung gestellt. Zentraler Aspekt der Übungen ist die Überführung theoretischer Analysekonzepte in praktische Implementierungen.

## **Zielsetzung und Zielgruppe**

Neben der Vermittlung von Kenntnissen über spezifische Verfahren und deren Anwendung liegt ein weiteres Ziel des Kurses darin, Möglichkeiten und Limitationen des Ansatzes aufzuzeigen und ihn so als wissenschaftliche Methode zu definieren. Die Inhalte des Kurses können auch als Erweiterung des Methodenkanons bzw. als neue Perspektive auf statistische Verfahren betrachtet werden, die in den Disziplinen der WVF eingesetzt werden.

Der Kurs könnte interessant für Sie sein, wenn Sie einen Einstieg in die Konzepte und Methoden des statistischen Lernens finden und/oder in die Lage kommen möchten, entsprechende Methoden selber anzuwenden.

Hinweis: Bei der Anmeldung können besondere inhaltliche Interessen vermerkt werden, die – unter Voraussetzung der Möglichkeit der zeitlichen und praktischen Umsetzung seitens des Dozenten – im Kurs mitberücksichtigt werden.

## **Ablauf des Workshops**

Der Kurs umfasst sechs Termine, von denen jeder einem ausgewählten Thema gewidmet ist.

### *Termin 1: Statistisches Lernen*

- Frage: Welche Fragen können mithilfe von Verfahren des statistischen Lernens gelöst werden? (Mögliche) Antwort: Unterscheidung zwischen Erklärung und Vorhersage mithilfe von Daten.
- Was ist die Grundidee des statistischen Lernens?  
Finden einer Funktion für die Relation zwischen Prädiktor- und Ergebnis-Variablen.
- Welche Verfahren des statistischen Lernens existieren?  
Supervidierte und nicht-supervidierte Verfahren.
- Welche Probleme treten beim Einsatz solcher Verfahren auf?  
Vorhersagegenauigkeit vs. Interpretierbarkeit von Modellen.
- Welche Unterschiede zu und Gemeinsamkeiten mit „klassischen“ (sog. frequentistischen) statistischen Verfahren gibt es?  
Auswahl des am besten geeigneten Vorhersagemodells vs. Test der Signifikanz eines beobachteten Unterschieds zweier experimenteller Bedingungen.
- Kurzdarstellung von R und RStudio

## Termin 2: Lineare Regression

- Schätzung der Parameter eines linearen Modells (Fitting mit der Least-Squares-Methode)
- Bedeutung einzelner Prädiktorvariablen und Kombinationen von Prädiktorvariablen
- Genauigkeit und Güte der Modellschätzung
- Vergleich linearer Regression mit k-Nearest-Neighbours-Verfahren
- Modellierung nicht-linearer Zusammenhänge zwischen Prädiktor- und Ergebnisvariablen, Link-Funktionen

## Termin 3: Klassifikationsverfahren

- Logistische Regression
- Lineare Diskriminanzanalyse

## Termin 4: Resampling-Methoden

- Kreuzvalidierungsverfahren
- Bootstrap

## Termin 5: Verbesserung linearer Modelle zur Selektion eines optimalen Modells

- Verwendung alternativer Fitting-Methoden
- Subset-Selektion
- Shrinkage-Verfahren (Ridge-Regression, LASSO-Regression)

## Termin 6: Nicht-supervidierte Verfahren

- Hauptkomponentenanalyse
- Clusteranalyse (k-Means-Clusters)