

Einführung in das maschinelle Lernen

Ziel

Die Teilnehmenden können nach dem Workshop die Grundlagen des statistischen maschinellen Lernens skizzieren, kennen die Voraussetzungen und verstehen die Herangehensweise verschiedener Methoden aus diesem Bereich bei unterschiedlichen Fragestellungen. Sie können diese Methoden computergestützt auf einen Datensatz anwenden.

Beschreibung

Zwar ist das lineare Modell mit seinen Generalisierungen in den empirischen Wissenschaften meist wohl bekannt, allerdings bringt es einen strukturellen Nachteil mit sich: Die*der Anwender*in muss bereits wissen, welcher grundlegende Zusammenhang zwischen erklärenden Variablen und zu erklärender Variable besteht. Erachtet man diesen Zusammenhang als gegeben, so schätzt man (meist mit der Methode der kleinsten Quadrate) die zugehörigen Parameter. Maschinelles Lernen hingegen nutzt überwiegend nichtparametrische Verfahren, also solche, die keine Annahme an die Art des Zusammenhangs mehr voraussetzen.

Simon (1983) definiert den maschinellen Lernvorgang als adaptive Änderungen eines Systems, in dem Sinne, dass sie das System in die Lage versetzen, die gleiche(n) Aufgabe(n) auf Basis der gleichen Population bei Wiederholung effizienter oder effektiver zu erfüllen.

Der Kurs geht auf die Grundlagen statistischen, maschinellen Lernens ein und diskutiert Hauptvertreter solcher Lernverfahren mit Schwerpunkt auf Klassifikationsmethoden (binär und multiclass).

Weitere Beispiele behandeln Regressionsmethoden sowie Verfahren des sogenannten unsupervised learnings. Eingebettet finden Übungen zu den vorgestellten Methoden mit dem Programmpaket R statt.

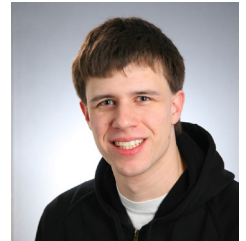
Voraussetzung

- Die Veranstaltung hat moderne statistische Verfahren zum Inhalt und setzt daher grundlegende Kenntnisse in Statistik (z. B. die Begriffe des Erwartungswertes oder der Standardabweichung sowie die Methode der kleinsten Quadrate) bereits voraus.
- Vorkenntnisse in R sind erforderlich.
- Die Teilnehmer*innen sollten R (<https://cran.r-project.org/>) auf ihren eigenen Computern installiert haben, um die Übungen bestreiten zu können. Details zu erforderlichen Paketen werden rechtzeitig bekannt gegeben.

Organizational Information

Sprache / Format	Deutsch / Präsenz
Zielgruppe	Promovierende in allen Phasen der Promotion und Postdocs (R2/R3) aller Fachbereiche
Datum	Montag-Mittwoch, 15.-17. Juli 2024, 9:00 – 16:00
Anmeldung	Für die Registrierung klicken Sie bitte hier

Workshopleitung



Dr. Florian Dumpert
Statistisches Bundesamt,
Wiesbaden

Referent im Bereich
„Künstliche Intelligenz, Big
Data“
(vormals: „Maschinelles
Lernen und Imputationsver-
fahren“)