

(Überwachtes) Maschinelles Lernen

Erarbeitet von
Dr. Katja Theune

Lernziele	1
Inhalt	2
Einstieg	2
Lerntypen, Problemtypen und Verfahren des maschinellen Lernens	2
Überwachtes Lernen	3
Abschluss	3
Weiterführendes Material	4
Disclaimer	4

Lernziele

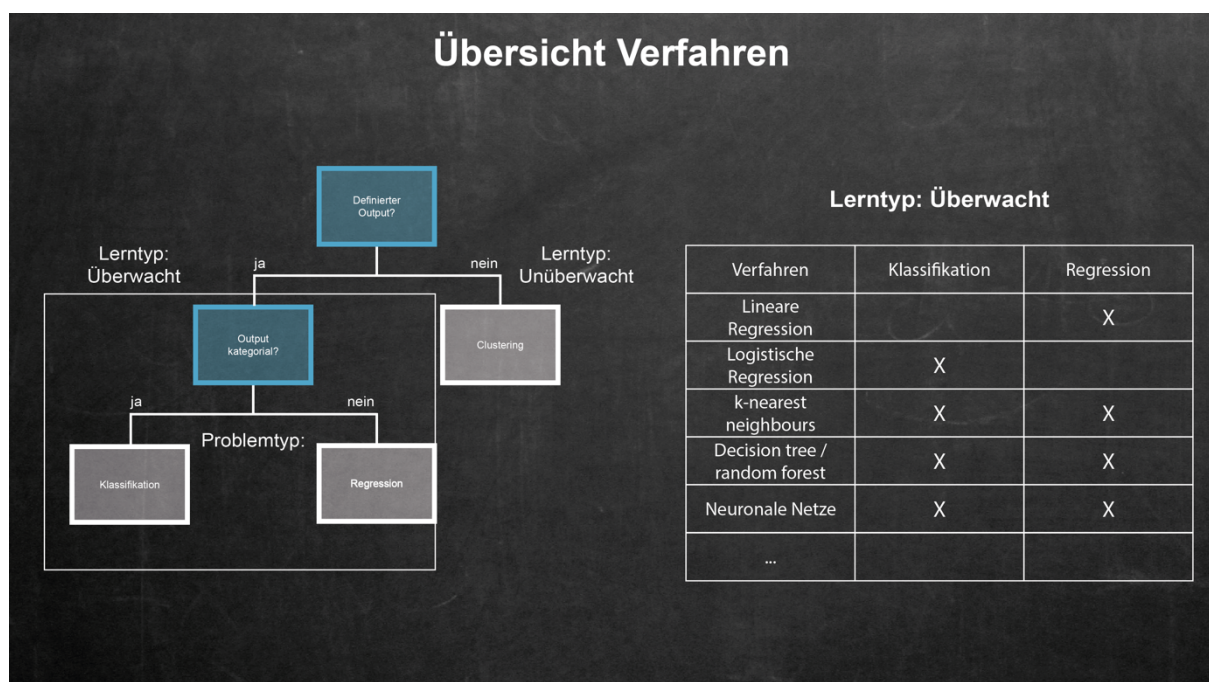
- Du kannst verschiedene Lerntypen des maschinellen Lernens nennen
- Du kannst verschiedene Problemtypen des maschinellen Lernens nennen
- Du kannst verschiedene Verfahren des überwachten maschinellen Lernens nennen
- Du kannst den groben Ablauf des überwachten Lernens erläutern

Inhalt

Einstieg

Wer sich über maschinelles Lernen informiert, wird schnell sehen, dass es sehr, sehr viele verschiedene Verfahren gibt, die sich in Einsatzmöglichkeiten und Arten des Lernens voneinander unterscheiden. Um einen Überblick über diese Verfahren zu bekommen, bietet sich zunächst an, sich mit den verschiedenen Lerntypen und Problemtypen zu beschäftigen. Der Fokus liegt hier aber auf überwachten Lernverfahren, da wir mit ihnen Prognosemodelle bilden wollen.

Lerntypen, Problemtypen und Verfahren des maschinellen Lernens



Wir unterscheiden zwischen drei großen Lerntypen, dem supervised bzw. überwachten Lernen, dem unsupervised bzw. unüberwachten Lernen und dem reinforcement bzw. bestärkenden Lernen. Letzteres betrachten wir hier aber nicht. Unüberwachtes Lernen verwenden wir üblicherweise zum Finden von Strukturen und uns vorher noch ganz unbekannten, undefinierten Gruppen in Daten. Das überwachte Lernen eignet sich insbesondere für Prognosen von vorher definierten bzw. gelabelten Outputs.

Der Problemtyp beschreibt quasi die Art unseres Outputs, also ob wir z. B. im überwachten Lernen mit einem kategorialen Output eine Klassifikation oder mit einem metrischen Output eine Regression anwenden müssen. Bei einer Klassifikation wollen wir für neue Beobachtungen bestimmte Klassen oder Gruppen prognostizieren. Ein Beispiel wäre hier die Zuordnung von Studierenden aufgrund bestimmter Eigenschaften zu den Klassen „Studienabbruch“ und „kein Studienabbruch“. Bei einer Regression wollen wir für eine neue

Beobachtung einen bestimmten Wert prognostizieren. Als Beispiel können wir uns die Prognose der Durchschnittsnote am Ende des Studiums vorstellen.
Beim unüberwachten Lernen begegnet uns dagegen als Problemtyp z. B. das Clustering.

Wir werden noch sehen, dass sich das jeweilige passende Lernverfahren für einen Anwendungsfall dann aus diesen beiden Überlegungen ergibt. Im Prinzip hängen Lerntypen, Problemtypen und die Verfahren alle miteinander zusammen bzw. bedingen sich gegenseitig. So lassen sich für bestimmte Lerntypen auch nur bestimmte Verfahren anwenden. Diese Verfahren sind aber wiederum nicht immer für alle Problemtypen geeignet.

Da wir Prognosemodelle aufstellen wollen, besprechen wir hier nur Verfahren des überwachten Lernens. Dazu gehört u. a. die lineare Regression, die logistische Regression, das k-nearest neighbours Verfahren, decision trees, random forests und neuronale Netze. Das sind natürlich nicht alle Verfahren, die es gibt, aber sie begegnen uns recht häufig in Forschung und Praxis. In der Tabelle ist angegeben, ob sie sich für eine Klassifikation, Regression oder beides eignen. Wir sehen, dass die meisten Verfahren für beide Problemtypen anwendbar sind.

Überwachtes Lernen

Für den weiteren Verlauf ist es hilfreich, dass wir uns nochmal kurz an die vereinfachte Idee des überwachten Lernens, die Konzepte des Trainings bzw. Lernens und der Evaluation erinnern. In einem ersten Schritt, dem Training, werden dem Lernverfahren Trainingsdaten zugeführt, die neben den Inputs bzw. Features auch einen definierten Output bzw. eine Zielgröße enthalten. Wir kennen also die zu den Inputs zugehörigen, richtigen Outputs und können das Verfahren somit überwachen. Mit diesen Input-Output-Paaren werden nun Gesetzmäßigkeiten bzw. Muster gelernt, um für bestimmte Inputs den, im besten Fall richtigen oder möglichst ähnlichen, Output vorherzusagen. Das Prinzip ist bei einer Regression und Klassifikation sehr ähnlich.

Nach dem Training eines Prognosemodells wird in einem zweiten Schritt das Modell evaluiert. D. h. es wird dann überprüft, wie gut die Prognose ist bzw. wie hoch der Fehler des Modells ist. Also z. B. wie häufig das Modell den Inputs den richtigen oder falschen Output zugeordnet hat. Wir vergleichen also den vom Modell prognostizierten Output mit dem wahren Output, den wir ja im überwachten Lernen kennen. Unser Ziel ist es, am Ende ein Prognosemodell zu erhalten, das möglichst „gute“ Prognosen macht, also für neue Beobachtung möglichst den passenden Output vorhersagt. Was genau „gut“ bedeutet, werden wir noch kennenlernen.

Abschluss

Wir haben uns die verschiedenen Lerntypen wie überwachtes und unüberwachtes Lernen und verschiedene Problemtypen wie Klassifikation und Regression angesehen. Aus diesen ergibt sich dann, welches der vielen Verfahren des maschinellen Lernens für unseren

Anwendungsfall überhaupt geeignet ist. Darüber hinaus haben wir uns nochmal daran erinnert, dass das überwachte Lernen aus Training und Evaluation besteht.

Weiterführendes Material

Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques* (3. Auflage). Morgan Kaufmann.

James, G., Witten, D., Hastie, T., & Tibshirani, R., & Tylor, J. (2023). *An Introduction to Statistical Learning - with Applications in Python*. Springer.

Lantz, B. (2015). *Machine learning with R* (2. Auflage). Packt Publishing Ltd, Birmingham.

Disclaimer

Transkript zu dem Video „Prognosemodelle (Klassifikation und Regression): (Überwachtes) Maschinelles Lernen“, Dr. Katja Theune.

Dieses Transkript wurde im Rahmen des Projekts ai4all des Heine Center for Artificial Intelligence and Data Science (HeiCAD) an der Heinrich-Heine-Universität Düsseldorf unter der Creative Commons Lizenz [CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/) veröffentlicht. Ausgenommen von der Lizenz sind die verwendeten Logos, alle in den Quellen ausgewiesenen Fremdmaterialien sowie alle als Quellen gekennzeichneten Elemente.