

Woche 04: Wie war das nochmal? – Überblick und Ausblick

Skript

Erarbeitet von
Dr. Maike Mayer

Lernziele	1
Inhalt	1
Einstieg	1
Begriffcheck – Was war was?	2
Zurück zu der Statistik	3
Auf geht's!	4
Quellen	5
Disclaimer	5

Lernziele

- Erinnern wichtiger Definitionen
- Nachvollziehen der Verbindungen zwischen Woche 3 und 4 sowie zwischen den Videos der vierten Woche

Inhalt

Einstieg

In Europa ziehen junge Erwachsene durchschnittlich mit 26,5 Jahren aus dem Elternhaus aus. In Deutschland im Schnitt hingegen schon mit 23,6 Jahren. Im Prüfungsjahr 2020 lag die mittlere Fachstudiendauer für einen Bachelorabschluss – hier anhand des Medians (ohne Lehramtsbachelor) – bei 7,2 Semestern für Erstabsolventinnen und Erstabsolventen.

Quelle [1, 2]

Einblendung Illustrationen

Statistische Angaben wie diese begegnen uns sehr häufig im Alltag. Es gibt Statistiken zu so gut wie allem. Und daran wird ein Stück weit deutlich: Statistik ist ein wichtiges Werkzeug, das uns dabei hilft, die Welt zu beschreiben, große Datenmengen zu strukturieren und neues Wissen daraus zu generieren. Daher ist es keine Überraschung, dass Statistik und statistische Konzepte auch für Künstliche Intelligenz und für Maschinelles Lernen wichtig sind.

Doch halt, da war es wieder ... Maschinelles Lernen ... Einer der Begriffe, den man im Zusammenhang mit Künstlicher Intelligenz so oft hört. In der letzten Woche hast du eine ganze Reihe von Begriffen kennengelernt, die einem häufig begegnen, wenn man sich mit Künstlicher Intelligenz beschäftigt. Da diese Begriffe auch für den weiteren Kurs sehr wichtig sind, werfen wir nochmal kurz einen Blick darauf.

Einblendung Schlagwort

Begriffsscheck – Was war was?

In der letzten Woche hast du Maschinelles Lernen (ML) als Teilgebiet der Künstlichen Intelligenz kennengelernt. Dieser Begriff ist besonders wichtig, da wir uns in diesem Kurs hauptsächlich auf Maschinelles Lernen konzentrieren. Maschinelles Lernen bedeutet – ganz grob gesagt – Lernen aus Daten. Es handelt sich um einen Prozess, bei dem beispielsweise ein Computer oder ein System selbstständig aus Daten lernt, ohne dass ein Mensch diesen Prozess Schritt für Schritt begleiten oder anleiten muss. Durch dieses Lernen verbessert der Computer dann eine Leistung oder präzisiert Vorhersagen.

Quelle [3, 4]

Methodisch betrachtet, lässt sich Maschinelles Lernen in insgesamt drei Kategorien einteilen. Bei der ersten Kategorie, dem Supervised Learning (also dem überwachten Lernen), gibt der Mensch vor, was gelernt werden soll, indem Trainingsdaten mit festen Labeln zur Verfügung gestellt werden. Anhand dieser Trainingsdaten lernt dann die Maschine, um später Vorhersagen treffen zu können.

Bei der zweiten Kategorie, dem Unsupervised Learning (also dem unüberwachten Lernen), wird im Vorfeld nicht vorgegeben, was gelernt werden soll. Die Maschine oder das System soll dann selbstständig Zusammenhänge und Strukturen in den Daten finden.

Und dann wäre da noch das Reinforcement Learning (also das verstärkende Lernen), bei dem ein sogenannter Agent oder eine sogenannte Agentin mit der Umgebung interagiert, um eine Belohnung zu maximieren.

Einblendung der Kategorien

Aber halt. Es gibt doch auch noch Deep Learning. Wie passt das denn jetzt da rein? Deep Learning (oder tiefes Lernen) ist ein Teilgebiet von Maschinellern Lernen. Für Deep Learning werden, wie der Name vielleicht schon vermuten lässt, große Datenmengen benötigt und die verwendeten Algorithmen sind vom menschlichen Gehirn und seinen Neuronen inspiriert. Das spiegelt sich in den Modellen wider, die typischerweise für Deep Learning verwendet werden, den Neuronalen Netzen. Hier werden mehrere Verarbeitungsebenen verwendet, um schrittweise immer detailliertere Merkmale aus Daten zu extrahieren. Mit Deep Learning als Teilgebiet des Maschinellen Lernens lassen sich ebenfalls Supervised Learning, Unsupervised Learning und Reinforcement Learning realisieren, und zwar mit Hilfe neuronaler Netze.

Quelle [3, 5]

Einblendung Schlagwörter

Deep Learning, Maschinellern Lernen und der Künstlichen Intelligenz als Überbegriff ist gemein, dass statistische Konzepte ein grundlegender Bestandteil für ihre Methoden sind ...

[Zurück zu der Statistik](#)

In der letzten Woche hast du viel über Trainingsdaten gelernt. Du weißt, dass man sie sich vorher sorgfältig anschauen sollte und dass man sich sorgfältig überlegen sollte, welche Daten und Angaben oder auch Features man für die Aufgabe, für die das System trainiert werden soll, benötigt. Du erinnerst dich vielleicht noch an das Beispiel mit dem Farn. Die Passung zwischen Daten und Aufgabe sowie die Qualität der Daten ist wichtig für die Genauigkeit des Modells. Wenn man mit Müll trainiert, kommt nämlich auch Müll raus. Oder im schlimmsten Fall eben diskriminierende Entscheidungen wie bei dem Amazon-Beispiel aus der vergangenen Woche, bei dem Frauen im Bewerbungsprozess aufgrund von verzerrten Trainingsdaten benachteiligt wurden.

Quelle [6]

Einblendung Illustrationen

Um Systemen, die auf Künstlicher Intelligenz basieren, gute Trainingsdaten zur Verfügung zu stellen, müssen diese Daten sorgfältig geprüft und vorverarbeitet werden. Hier können uns statistische Kennzahlen und Konzepte helfen, indem sie uns dabei unterstützen, die Trainingsdaten zu verstehen, aufzubereiten, zu visualisieren und zu interpretieren. Daher lernst du in dieser Woche einige wichtige statistische Konzepte und Kennzahlen, wie beispielsweise Matrizen, den Median oder auch einen Mittelwert, kennen. Und wir zeigen dir ein paar Fallstricke, auf die man bei der praktischen Anwendung dieser Konzepte und

Kennzahlen achten sollte: Es ist wichtig zu wissen, welche statistischen Kennzahlen man für seine Daten sinnvoll einsetzen kann. Und dafür sollte man sich seiner Daten bewusst sein. Nicht, dass man am Ende einen Gorilla übersieht. Was der Gorilla damit zu tun hat? Das erfährst du selbstverständlich auch in dieser Woche!

Einblendung Schlagwörter/Illustrationen

Aber wir zeigen dir natürlich auch – um nochmal zurück zum Thema zu kommen –, wie du dir beispielsweise einen Mittelwert eines Datensatzes mit Python ausgeben lassen kannst. Dafür benötigt Python sogenannte Module. Module sind im Prinzip eine Art Werkzeugkasten mit verschiedenen Hilfsmitteln, die bereits programmiert wurden. Wir stellen dir diese Woche zwei solcher Module vor, nämlich das Modul Scientific Python (SciPy) und Numerical Python (NumPy). Das Modul NumPy enthält auch eine neue Datenstruktur, nämlich numpy-array. Diese Datenstruktur ähnelt den verschachtelten Listen, die du bereits kennst. Und auch numpy-arrays wollen wir dir diese Woche näher zeigen. Du lernst, wie man sie erzeugt und auch, wie man auf einzelne Elemente eines Arrays zugreifen oder diese ändern kann.

Einblendung Schlagwörter/Illustrationen

Dafür brauchst du das Wissen zu Listen aus der letzten Woche. Besonders wichtig: die Indices. Damit kannst du Elemente einer Liste, aber auch eines Arrays ansprechen. Denk daran: Geht man beispielsweise eine Liste von links nach rechts durch, hat das erste Element den Index 0, das nächste den Index 1 und so weiter. Geht man die Liste von rechts nach links durch, hat das erste Element den Index -1, das zweite Element den Index -2 und so weiter.

Einblendung Indices

Auf geht's!

In dieser Woche lernst du also einige statistische Konzepte und Kennzahlen wie Mittelwerte und Matrizen kennen. Wir zeigen dir anhand von SciPy und NumPy Module und Funktionen in Python. Mit diesen Modulen kannst du dann unter anderem auch die statistischen Kennzahlen, die du in dieser Woche kennenlernst, mit Python ermitteln. Im Bereich der Ethik wird es in dieser Woche um Nachhaltigkeit gehen: Du erfährst, wie KI-Systeme Nachhaltigkeit unterstützen können, beispielsweise bei der Optimierung von Energieverbrauch. Aber wir wollen auch über die Nachhaltigkeit von KI-Anwendungen selbst sprechen. Abschließend zeigen wir dir noch ein paar Beispiele, worauf man bei der praktischen Anwendung von statistischen Kennzahlen und Konzepten achten sollte und warum es wichtig ist, seine Daten zu kennen.

Einblendung der Videotitel

Nun aber viel Spaß in dieser Woche! Und schau doch mal, ob du den Gorilla findest.

Quellen

- Quelle [1] Statistisches Bundesamt (2023). *Jugend in Zahlen*. Abgerufen unter <https://www.destatis.de/DE/Im-Fokus/Jahr-der-Jugend/inhalt.html>, letzter Zugriff: 3. Februar 2023.
- Quelle [1] Statistisches Bundesamt. *Bildung in Zahlen 2020/2021 im Digitalen Magazin*. Abgerufen unter: <https://www.destatis.de/DE/Mediathek/Digitales-Magazin/Bildung/inhalt.html#579808>, letzter Zugriff: 3. Februar 2023.
- Quelle [3] Cambridge Dictionary
- Quelle [4] Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2018). *Foundations of machine learning*. MIT press.
- Quelle [5] Oxford Languages
- Quelle [6] Hamilton, I. A. (2018, October 11). Amazon hat eine Künstliche Intelligenz für Bewerbungen gebaut, die Frauen diskriminierte. Business Insider. <https://www.businessinsider.de/wirtschaft/amazon-hat-eine-ki-gebaut-die-frauen-diskriminierte-2018-10/>, letzter Zugriff: 3. Februar 2023.

Disclaimer

Transkript zu dem Video „Woche 04: Wie war das nochmal? – Überblick und Ausblick“, Dr. Maike Mayer.

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.