

Woche 08: Wie war das nochmal? – Überblick und Ausblick

Skript

Erarbeitet von
Dr. Maike Mayer

Lernziele	1
Inhalt	1
Einstieg	1
Den Wald vor lauter Bäumen nicht sehen?	2
Seinen Augen nicht trauen	3
Auf geht's!	3
Quellen	4
Disclaimer	4

Lernziele

- Erinnern wichtiger Begriffe
- Nachvollziehen der Zusammenhänge zwischen den Inhalten

Inhalt

Einstieg

Kennst du auch diese wilden Persönlichkeitstests, die wie eine große Mindmap aufgebaut sind und bei denen man, je nach eigener Antwort, einen anderen Pfad entlang gehen muss? Im Prinzip versuchen diese Tests – wissenschaftlich in der Regel nicht wirklich fundiert – dich einer Klasse oder Kategorie zuzuordnen.

In der letzten Woche hast du einen Ansatz für Klassifikationen kennengelernt, der auf Maschinellem Lernen basiert, den k-nearest neighbours Algorithmus. Mit Hilfe dieses

Algorithmen lassen sich neue Beobachtungen einer Klasse zuordnen. Dafür wird eine bestimmte Anzahl an Beobachtungen einbezogen, die am nächsten an der neuen Beobachtung liegen. In dieser Woche lernst du ein weiteres Verfahren kennen, mit dem man Beobachtungen klassifizieren kann: die sogenannten Entscheidungsbäume. Auch hier handelt es sich um eine statistische Lernmethode des Maschinellen Lernens. Diese Bäume lassen sich sogar zu Entscheidungswäldern erweitern ...

Einblendung Illustrationen

Den Wald vor lauter Bäumen nicht sehen?

Bei Entscheidungsbäumen handelt es sich um baumbasierte Klassifikationsverfahren aus dem Bereich des Supervised Learnings. Diese Kategorie des Maschinellen Lernens kennst du mittlerweile schon gut – hier gibt der Mensch vor, was gelernt werden soll! Diese Bäume können jedoch nicht nur für Klassifikationen, sondern auch für Regressionen, die du schon kennst, eingesetzt werden.

Tatsächlich ist der Name „Entscheidungsbaum“ schon sehr beschreibend. Du kannst dir dieses Verfahren tatsächlich wie eine Art Baum vorstellen, der auf dem Kopf steht. Die Trainingsdaten bilden die Wurzeln und die Blätter sind die Klassen bzw. Kategorien, in die die Beobachtungen eingeteilt werden. Mit Hilfe von verschiedenen Merkmalen werden dann Stück für Stück Entscheidungen getroffen, in welche Kategorie bestimmte Beobachtungen gehören. Das Ziel dabei ist, die Klassen möglichst gut zu trennen – also in einem Blatt möglichst viele Beobachtungen einer Klasse zu haben. Das klingt vielleicht noch etwas abstrakt, aber keine Sorge, wir erklären dir dieses Verfahren in dieser Woche näher.

Einblendung Schlagwörter/Illustrationen

Man kann übrigens statt eines einzelnen Baums auch für die Klassifikation von Beobachtungen einen ganzen Wald heranziehen. Kein Witz! Das nennt man dann einen Entscheidungswald oder Decision Forest. Eine Variante des Entscheidungswalds stellen wir dir ebenfalls in dieser Woche vor, nämlich den Random Forest. Jeder Baum dieses Waldes wird auf Basis einer unterschiedlichen Teilmenge des Trainingsdatensatzes erstellt – es werden also unterschiedliche Beobachtungen und auch unterschiedliche Merkmale für die Klassifikation herangezogen. Die finale Klassenzugehörigkeit einer neuen Beobachtung hängt dann davon ab, in welche Klasse die Beobachtung von den meisten Bäumen des Waldes einsortiert wurde.

Einblendung Schlagwörter/Illustrationen

Damit diese Bäume und Wälder nicht nur Theorie bleiben, zeigen wir dir in Python auch, wie du Entscheidungsbaummodelle und Entscheidungswälder erstellen kannst. Dafür wird das Modul Scikit-learn, die Datenstruktur Numpy Array und auch das Qualitätsmaß der Genauigkeit bzw. Accuracy (also wie viel Prozent der gemachten Klassifikationen korrekt

waren) wieder relevant. Und auch über die Überanpassung bzw. Overfitting eines Modells und über Outlier werden wir in dieser Woche nochmal sprechen. Als kleine Gedächtnisauffrischung: Bei einer Überanpassung wird der Zusammenhang zwischen Features und Zielmerkmal überschätzt. Das Modell wird zu stark an die Trainingsdaten angepasst, was zwar zu einem kleinen Fehler während des Trainings führt, jedoch mit einem großen Fehler beim Testen einhergeht, da das Modell nicht gut auf die Testdaten passt. Bei Ausreißern wiederum handelt es sich um Datenobjekte, bei denen sich der Wert eines Features oder einer Kombination von Features stark von dem typischen Wert der anderen Datenobjekte unterscheidet.

Einblendung Schlagwörter/Illustrationen

Mit diesem Wissen und dem Wissen aus den vergangenen Wochen wirst du den Wald auf jeden Fall trotz der ganzen Bäume sehen! Aber apropos Sehen ...

Seinen Augen nicht trauen

Um Sehen wird es ebenfalls in dieser Woche gehen, und zwar im Bereich der Ethik, denn wir schauen uns in dieser Woche sogenannte Deepfakes an. Mit diesem Begriff werden meist realistisch wirkende Medieninhalte bezeichnet, die aber durch KI-basierte Techniken abgeändert und damit verfälscht worden sind. Es ist lustig, wenn die Mona Lisa plötzlich Grimassen schneidet, aber es ist nicht mehr so lustig oder sogar plötzlich extrem ernst, wenn überzeugend damit Fehlinformationen verbreitet werden oder man vielleicht sogar selbst Opfer eines Deepfakes geworden ist.

Quelle [1]

Deepfakes basieren auf Deep Learning und bei Deepfakes kann man Stimmen ändern, Mimik oder Gestik generieren oder sogar ganze Personen künstlich erzeugen. Wir zeigen dir in dieser Woche verschiedene Varianten des Deepfakes wie synthetisch erstellte Avatare, Face Swapping oder Face Reenactment. Manchmal muss man da wirklich mehr als einmal hinschauen und nicht immer kann man seinen Augen trauen.

Auf geht's!

In dieser Woche lernst du also Entscheidungsbäume und Entscheidungswälder kennen. Wir erläutern dir zunächst die Theorie hinter diesen Verfahren und zeigen dir dann, wie du Entscheidungsbaummodelle und Entscheidungswälder in Python erzeugen kannst. Außerdem sprechen wir diese Woche über Deepfakes und über die ethischen Konsequenzen bzw. Herausforderungen, die mit ihnen einhergehen. Und wir haben auch noch ein Anwendungsbeispiel für Entscheidungsbäume für dich vorbereitet, nämlich ein Forschungsprojekt an der Heinrich-Heine-Universität Düsseldorf.

Einblendung Videotitel

Jetzt aber erstmal Augen auf und viel Spaß in Woche 8!

Quellen

Quelle [1] Zsolnai-Fehér, K. (2020, April 15). *Sure, DeepFake Detectors Exist—But Can They Be Fooled?* [Video].

Disclaimer

Transkript zu dem Video „Woche 08: Wie war das nochmal? – Überblick und Ausblick“, Dr. Maïke 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.