

Arbeitsblatt: Scheinkorrelationen

Kurzbeschreibung

Korrelationen sollten mit Vorsicht genossen werden! Vor allem, wenn wenige Datenpunkte vorliegen, findet man häufig zufällige Korrelationen. In diesem Arbeitsblatt wird ein Beispiel vorgestellt, anschließend sehen Sie sich in Gruppen weitere Beispiele an.

Vorbereitung

Sehen Sie sich die im Moodle-Kurs verlinkten Videos zu Korrelation an. (Es handelt sich um folgende Videos:)

https://youtu.be/B6qRO661_ek

<https://youtu.be/y5DiAVHYzCs>

<https://youtu.be/WilIXlouYWc>

<https://youtu.be/NSXylf15ps4>

Aufgabe

Nicht alle Korrelationen, die man in Daten findet, entstehen aufgrund „echter“ Zusammenhänge. Man spricht dann von „Scheinkorrelationen“. Ein beliebtes Beispiel für eine Scheinkorrelation ist eine Korrelation, die

hier

<http://opentextbc.ca/researchmethods/wp-content/uploads/sites/37/2015/09/chart.png>

dargestellt wird: eine Korrelation zwischen der Menge von US-Amerikaner*innen, die pro Jahr nach einem Sturz in den Swimming Pool ertrunken sind, und der Anzahl von Filmen pro Jahr, in denen Nicolas Cage mitgewirkt hat. Es dürfte klar sein, dass ein echter Zusammenhang zwischen den beiden Variablen (*Fälle von Ertrinken im Pool* und *Filme mit Cage*) nicht plausibel ist. Sieht man sich die Grafik an, sieht man, dass die Zahlen nur für die Jahre 1999 bis 2009 vorliegen - es ist sehr wahrscheinlich, dass diese Korrelation nicht auftreten würde, wenn man stattdessen Zahlen für andere Jahre vorliegen hätte. Falls die Grafik unter der o.g. URL nicht mehr existiert, können Sie sie bestimmt durch Suchen im Internet woanders finden.

Solche Fälle von Korrelation bezeichnet man auch als *Scheinkorrelation*. Der Begriff ist etwas irreführend - denn die Korrelation besteht ja tatsächlich - sie ist aber zufällig entstanden und würde für neue Daten sicher nicht wieder gefunden werden.

Es gibt eine weitere Art von Scheinkorrelationen: solche, bei denen die beiden Faktoren deshalb miteinander korrelieren, weil sie beide durch einen weiteren, dritten Faktor beeinflusst werden. Z.B. korreliert der Sachschaden, der durch ein Feuer entsteht, mit der Anzahl an Feuerwehrleuten, die zum Löschen geschickt wurden - aber nicht, weil mehr Feuerwehrleute mehr Sachschaden verursachen, sondern, weil beides von der Größe des

Arbeitsblatt „Arbeitsblatt: Scheinkorrelationen“ von Antje Schweitzer / KI B3 – KI in die berufliche Bildung bringen. Gefördert als InnoVET-Projekt aus Mitteln des Bundesministeriums für Bildung und Forschung.



Lizenziert unter CC BY 4.0 (<https://creativecommons.org/licenses/by/4.0/deed.de>)

Feuers abhängt: Je größer das Feuer, desto mehr Sachschaden entsteht, und desto mehr Feuerwehrleute werden zum Löschen benötigt. Auch hier spricht man von Scheinkorrelation zwischen Anzahl der Feuerwehrleute und Höhe des Sachschadens - und meint damit, dass es sich nicht um einen *ursächlichen* Zusammenhang handelt. Dieser besteht nur zur Größe des Feuers. Die Benennung ist etwas verwirrend, weil eine Korrelation sehr wohl besteht und auch dann zu erwarten ist, wenn man neue Daten zu anderen Feuern bekommt.

Zufällige Scheinkorrelationen sind für Maschinelles Lernen problematisch: ein Modell, das die Anzahl der Tode durch Ertrinken mithilfe der Anzahl der Filme vorhersagt, in denen Nicolas Cage mitspielt, ist zwar auf den oben genannten Daten erfolgreich, aber bei der Vorhersage neuer Daten sicher erfolglos.

Scheinkorrelationen aufgrund von anderen Faktoren sind dagegen für Maschinelles Lernen erwünscht: wenn der Sachschaden, der bei einem Feuer entstanden sein könnte, vorhergesagt werden soll, könnte es durchaus hilfreich sein, zu wissen, wie viele Leute zum Löschen benötigt wurden - die Ursächlichkeit ist hier nicht wichtig.

Die folgenden Links sind Quellen für weitere kuriose (und zufällige) Scheinkorrelationen. Bilden Sie Kleingruppen. Suchen Sie sich ein Beispiel aus und stellen Sie es in der Kleingruppe vor. Erwähnen Sie dabei auch, auf wie vielen Datenpunkten die Scheinkorrelation in diesem Fall basiert (falls diese Information vorhanden ist).

<https://www.eduacademy.at/gwb/mod/page/view.php?id=5195>

https://rp-online.de/panorama/wissen/lustige-scheinkorrelationen_bid-19620739

<https://scheinkorrelation.jimdofree.com>

Anmerkung: Sollten die Quellen nicht mehr existieren, kann man durch Suchen im Internet problemlos andere Beispiele für Scheinkorrelationen finden.