

Handbuch: 7.1.2. Modellierungsprozess

Um vom Datensatz zu einem fertigen Modell zu gelangen, bedarf es mehrerer Schritte. Einige dieser Schritte erfolgen automatisch durch die Software und bleiben für den Nutzer unsichtbar, aber wir wollen sie zum besseren Verständnis hier beschreiben.

Als Erstes müssen wir eine Formel bzw. Gleichung wählen, die am besten in der Lage ist, die uns zur Verfügung stehenden Daten zu repräsentieren. Haben wir es beispielsweise mit einem idealen Gas zu tun, dann wird die Zustandsgleichung idealer Gase eine gute Wahl sein. Ist das Gas nicht ideal, wird diese Gleichung nur in Annäherung funktionieren. Wir müssen uns der der Gleichung innewohnenden Begrenzungen bewusst sein, wenn wir anfangs die Gleichung auswählen. Es ist an dieser Stelle gut, sich auch bewusst zu machen, dass neuronale Netze erwiesenermaßen in der Lage sind, jeden Datensatz mit großer Genauigkeit darzustellen. Aus diesem Grund verwenden wir neuronale Netze als die Gleichung unserer Wahl.

Als Zweites wählen wir die Daten aus, mit denen das Modell trainiert werden soll. Damit das überhaupt Sinn macht, müssen wir verstehen, dass die Parameter eines mathematischen Modells zu diesem Zeitpunkt noch nicht bekannt sind. Um noch einmal auf die Gleichung für ideale Gase zurückzugreifen: Das hieße, dass wir den Wert der idealen Gaskonstanten R nicht kennen würden. Um den Wert des Parameters zu bestimmen, benötigen wir die Daten für alle Variablen dieser Gleichung. Wollen wir den besten Wert des Parameters (oder der Parameter) der Gleichung aufgrund der historischen Daten für diese Variablen bestimmen, so sprechen wir davon, das Modell zu trainieren.

Als Drittes überprüfen wir unsere Arbeit. Es ist üblich, den Datensatz in zwei Teile zu unterteilen. Einer wird verwendet, um das Modell zu trainieren. Der andere wird verwendet, um zu überprüfen, ob das Modell auch für solche Daten sinnvolle Ergebnisse erzielt, die nicht fürs Training verwendet wurden. Hier sprechen wir von Validierungsdaten. Ergeben die zuvor bestimmten Parameterwerte genaue Ergebnisse für die Validierungsdaten, ist das Modell gut und wir können den Modellierungsvorgang als beendet betrachten.