

Handbuch: 7.1.3. Die unabhängigen Variablen auswählen

Wenn wir noch einmal auf die Zustandsgleichung für ideale Gase zurückblicken, so können wir sehen, dass wir drei unabhängige Variablen haben, nämlich den Druck P , das Volumen V , und die Temperatur T , um daraus die Gasmenge n zu berechnen. Das erfordert das Modell. Aber nehmen wir einmal für den Augenblick an, wir wüssten das nicht.

Der Datensatz wird viele Variablen enthalten, die uns zur Verfügung stehen. Da gibt es wahrscheinlich verschiedene Drücke und Temperaturen. Es mag auch Vibrationen, Flüsse, Ebenen, Energieverbrauch usw. geben. Aus dieser Vielfalt von Angeboten müssen wir die drei unabhängigen Variablen auswählen, die uns die gewünschte Informationen liefern, d.h. jene Informationen, die wir brauchen, um die für uns interessante abhängige Variable zu berechnen.

Es gibt zwei Arten, diese Auswahl zu treffen. Als Erstes könnten wir unsere menschliche Intuition und unseren Sachverstand einsetzen. Es waren solche Einsichten, die beim idealen Gas zur Geltung gebracht wurden. Dieser Ansatz erfordert Sachverstand und auch etwas Zeit, ihn umzusetzen. Weil unser Sachverstand aber nicht in allen Fällen endgültig und eindeutig ist, sondern vielmehr hypothetischen Charakter hat, bedarf es mehrerer Versuchsschleifen von Versuch und Irrtum, bis wir schließlich ein gutes Modell gefunden haben. Das ist freilich ein aufwändiger Prozess.

Als Alternative könnten wir den Computer einsetzen, um für uns eine Datenanalyse durchzuführen. Es gehört zu den wichtigen Vorzügen dieser Software, dass die Auswahl von unabhängigen Variablen automatisch erfolgt und dass mit hoher Wahrscheinlichkeit ein Satz von Variablen selektiert wird, die ein gutes Modell ergeben.

Auf der Bearbeitungsseite für die dynamische Grenzwerte oder für Soft-Sensoren finden Sie unter der Überschrift "Unabhängige Variablen auswählen" einen Knopf zur automatischen Auswahl unabhängiger Variabler. Diese Funktion geht davon aus, dass Sie bereits die Trainingszeiten und die Ausschlussbedingungen definiert haben. Sie wird sich dann die Daten holen, die mit diesen Vorbedingungen übereinstimmen und dann eine Korrelations-Analyse zwischen der abhängigen Variable und allen anderen Variablen wie auch eine Analyse aller Variablen untereinander durchführen. Die Analyse wird dann solche Variablen auswählen, die eine hohe Korrelation mit der abhängigen Variable, aber wenig Korrelation miteinander haben. Dadurch wird sichergestellt, dass wir ausreichende, aber keine redundanten Informationen im Datensatz haben, weil wir ja die Zahl der unabhängigen Variablen so niedrig wie möglich halten wollen, ohne dadurch die Genauigkeit des Modells zu gefährden. Nach der automatischen Auswahl können Sie die Liste der selektierten Variablen, so gewünscht, gerne noch nachbearbeiten.

Nachdem nun die unabhängigen Variablen ausgewählt wurden, kann das Modell trainiert werden.

