

Data Sciences Multivariate Analysen

Prof. Dr. Martin Przewloka
Institut für Digitale Assistenzsysteme e.V.
martin.przewloka@institut-das.de

Fokus: Python Framework

PCA - Beispielfolien
© www.institut-das.de

Multivariate Analysen - was ist das?

Bis jetzt: 2 Variablen/Merkmale

In der Praxis/Realität: viele Variablen/Merkmale

Beispiele:

Verkehrstauvorhersage (-wahrscheinlichkeit):
Wochentag, Datum, Uhrzeit, Wetter, Veranstaltungen, uvm.

Erwartete Restlebensdauer einer technischen Komponente:
Gesamtbetriebsstunden, Betriebsart, Lastprofil,
Wartungshistorie uvm.

Kernfragen:

- a) Wie ist der geschätzte Zusammenhang der Merkmale zum Ergebnis?
- b) Welche Merkmale zeigen den stärksten Zusammenhang, welche den schwächsten? (Relevanz)
- c) Kann ich Vorhersagen treffen wie bspw. zur Restlebensdauer einer Maschine?

Multiple Regression - Praxis

Regression ausführen:

4.) Fit berechnen

```
reg.fit(dataframe[['Family', 'Generosity', 'Freedom',  
'Health..Life.Expectancy']], dataframe['Happiness.Score'])
```

```
Family ... Life.Expectancy = Regressoren  
Happiness.Score = Regressand
```

5.) Ergebnis ausgeben:

```
reg.coef_
```

```
liefert gerundet: array([1.402 , 0.253 , 1.828 , 2.278])
```

```
reg.intercept_
```

```
liefert gerundet: 1.622
```

Interpretation?

Extrapolation - Vorhersagen

Ein wesentliches Ziel der Regressionsanalyse ist die Vorhersage (Prädiktion).

Dies kann (wenn wir der Einfachheit halber Zeitreihen unterstellen):

- Vergangenheitswerte
- Zwischenwerte
- Zukunftswerte

betreffen.

In unserem Fall wollen wir folgende Fragen beantworten:

- a) Welchen Wert für die Happiness würden wir für ein Land mit Family = 1.3, Generosity = 0.3, Freedom = 0.5 und Life.Expectancy = 0.7 erwarten?
- b) Wie würde sich die Happiness verschieben, wenn wir bei sonst gleichen Daten Freedom auf 0.6 anheben könnten?