

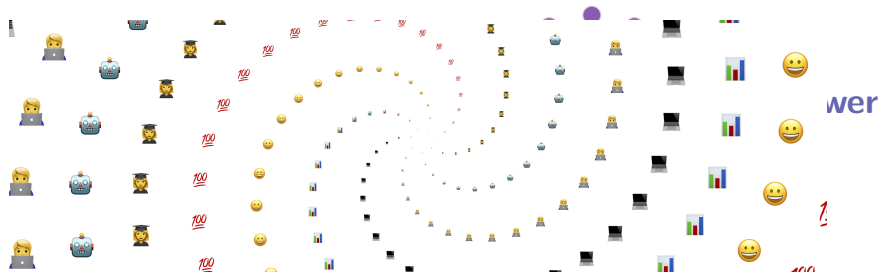
Daten Vorhersagen

Lineare Regression

Naoki Peter¹ Cyril Blum²

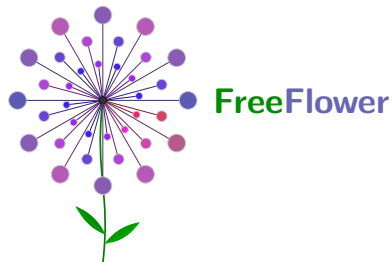
¹Fachschaft Informatik
Kantonsschule Zürich-Nord

²Fachschaft Informatik
Kantonsschule im Lee

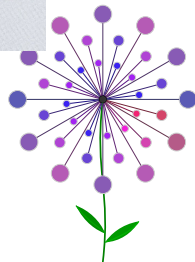
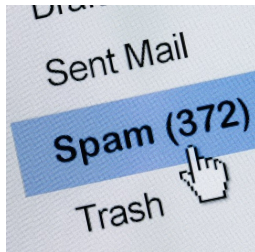


Themenübersicht

- Daten verstehen (analysieren und visualisieren)
- Zahlen vorhersagen**
- Kategorien vorhersagen

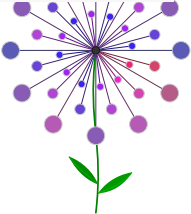
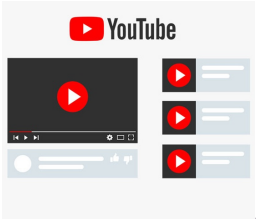
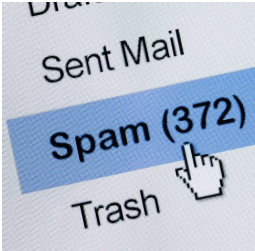


Maschinelles Lernen



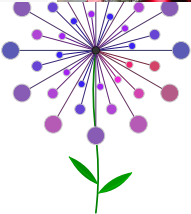
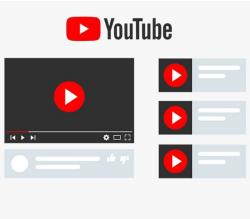
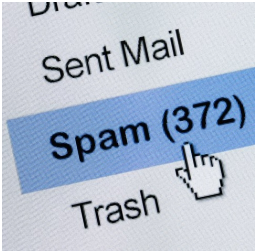
FreeFlower

Maschinelles Lernen



FreeFlower

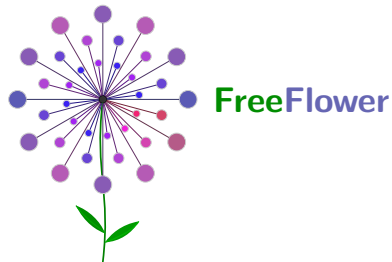
Maschinelles Lernen



FreeFlower

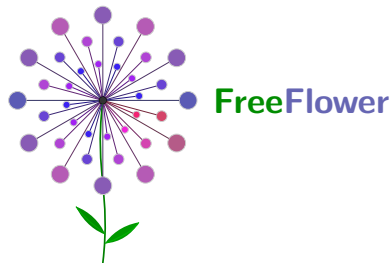
Lernziele

- Ich kann erklären, was eine **lineare Regression** ist und wofür sie verwendet wird.



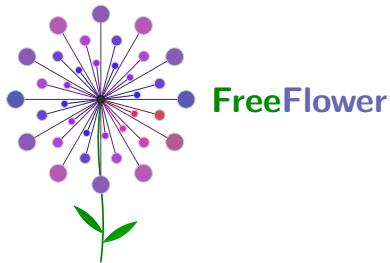
Lernziele

- Ich kann erklären, was eine **lineare Regression** ist und wofür sie verwendet wird.
- Ich kann den Unterschied zwischen **unabhängiger** und **abhängiger** Variable erläutern.



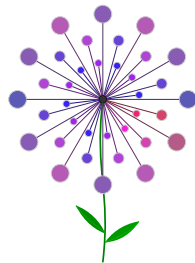
Lernziele

- Ich kann erklären, was eine **lineare Regression** ist und wofür sie verwendet wird.
- Ich kann den Unterschied zwischen **unabhängiger** und **abhängiger** Variable erläutern.
- Ich kann eine **lineare Funktion** aus zwei Punkten bestimmen.



Lernziele

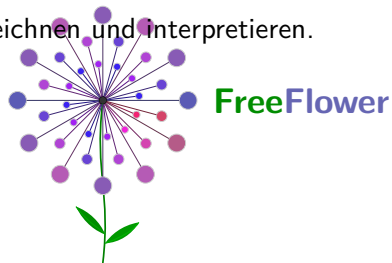
- Ich kann erklären, was eine **lineare Regression** ist und wofür sie verwendet wird.
- Ich kann den Unterschied zwischen **unabhängiger** und **abhängiger** Variable erläutern.
- Ich kann eine **lineare Funktion** aus zwei Punkten bestimmen.
- Ich kann das Prinzip der **Kleinste-Quadrate-Methode (OLS)** erklären und anwenden.



FreeFlower

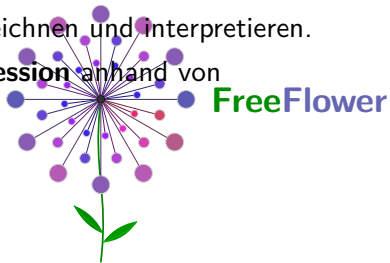
Lernziele

- Ich kann erklären, was eine **lineare Regression** ist und wofür sie verwendet wird.
- Ich kann den Unterschied zwischen **unabhängiger** und **abhängiger** Variable erläutern.
- Ich kann eine **lineare Funktion** aus zwei Punkten bestimmen.
- Ich kann das Prinzip der **Kleinste-Quadrate-Methode (OLS)** erklären und anwenden.
- Ich kann eine Regressionsgerade zeichnen und interpretieren.



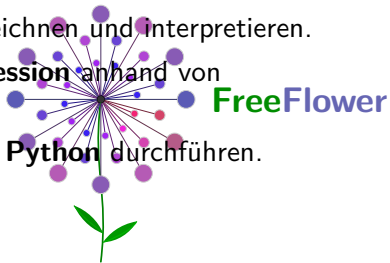
Lernziele

- Ich kann erklären, was eine **lineare Regression** ist und wofür sie verwendet wird.
- Ich kann den Unterschied zwischen **unabhängiger** und **abhängiger** Variable erläutern.
- Ich kann eine **lineare Funktion** aus zwei Punkten bestimmen.
- Ich kann das Prinzip der **Kleinste-Quadrate-Methode (OLS)** erklären und anwenden.
- Ich kann eine Regressionsgerade zeichnen und interpretieren.
- Ich kann die **Qualität einer Regression** anhand von Fehlermassen beurteilen.



Lernziele

- Ich kann erklären, was eine **lineare Regression** ist und wofür sie verwendet wird.
- Ich kann den Unterschied zwischen **unabhängiger** und **abhängiger** Variable erläutern.
- Ich kann eine **lineare Funktion** aus zwei Punkten bestimmen.
- Ich kann das Prinzip der **Kleinste-Quadrate-Methode (OLS)** erklären und anwenden.
- Ich kann eine Regressionsgerade zeichnen und interpretieren.
- Ich kann die **Qualität einer Regression** anhand von Fehlermassen beurteilen.
- Ich kann eine lineare Regression in **Python** durchführen.

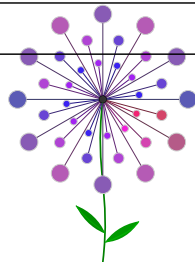


Zahlen vorhersagen mit Regression

Unabhängige und abhängige Variablen

Unabhängige Variable

Abhängige Variable



FreeFlower

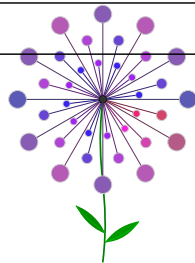
Zahlen vorhersagen mit Regression

Unabhängige und abhängige Variablen

Unabhängige Variable

Abhängige Variable

Lernzeit → Note in der Prüfung

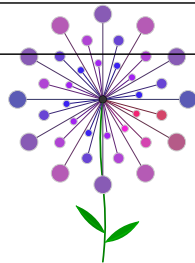


FreeFlower

Zahlen vorhersagen mit Regression

Unabhängige und abhängige Variablen

Unabhängige Variable		Abhängige Variable
Lernzeit	→	Note in der Prüfung
Alter eines Autos	→	Wiederverkaufswert

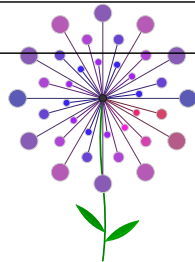


FreeFlower

Zahlen vorhersagen mit Regression

Unabhängige und abhängige Variablen

Unabhängige Variable		Abhängige Variable
Lernzeit	→	Note in der Prüfung
Alter eines Autos	→	Wiederverkaufswert
Anzahl Schritte pro Tag	→	Kalorienverbrauch

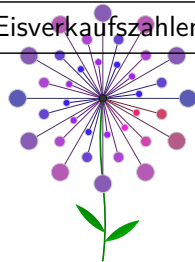


FreeFlower

Zahlen vorhersagen mit Regression

Unabhängige und abhängige Variablen

Unabhängige Variable		Abhängige Variable
Lernzeit	→	Note in der Prüfung
Alter eines Autos	→	Wiederverkaufswert
Anzahl Schritte pro Tag	→	Kalorienverbrauch
Ausstemperatur	→	Eisverkaufszahlen

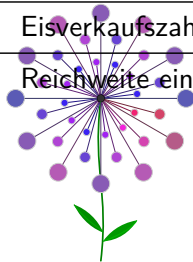


FreeFlower

Zahlen vorhersagen mit Regression

Unabhängige und abhängige Variablen

Unabhängige Variable		Abhängige Variable
	Lernzeit →	Note in der Prüfung
	Alter eines Autos →	Wiederverkaufswert
	Anzahl Schritte pro Tag →	Kalorienverbrauch
	Ausstemperatur →	Eisverkaufszahlen
	Anzahl Likes auf Social Media →	Reichweite eines Posts



FreeFlower

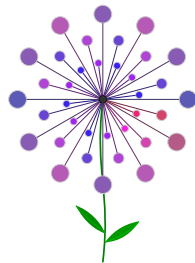
Unabhängige und abhängige Variablen

Unabhängige Variable

Kontrollierbare Grösse, beeinflusst andere Variable(n).

Abhängige Variable

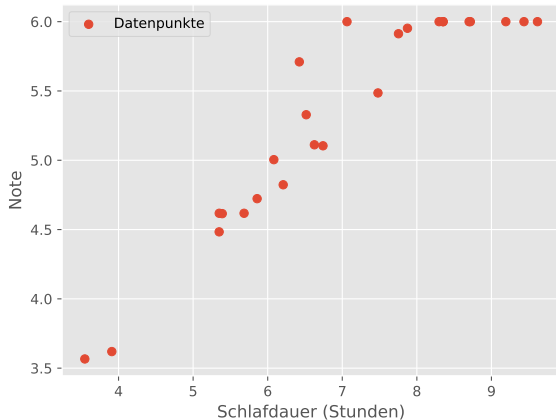
Messbare Grösse, wird beeinflusst von *unabhängiger* Variable.



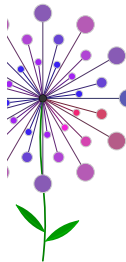
FreeFlower

Zahlen vorhersagen mit Regression

Unabhängige und abhängige Variablen



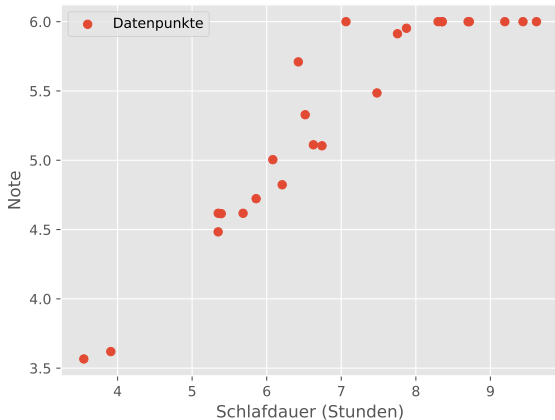
► **Unabhängige Variable:**
Schlafdauer



FreeFlower

Zahlen vorhersagen mit Regression

Unabhängige und abhängige Variablen



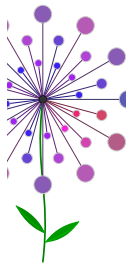
▶ **Unabhängige**

Variable:

Schlafdauer

▶ **Abhängige**

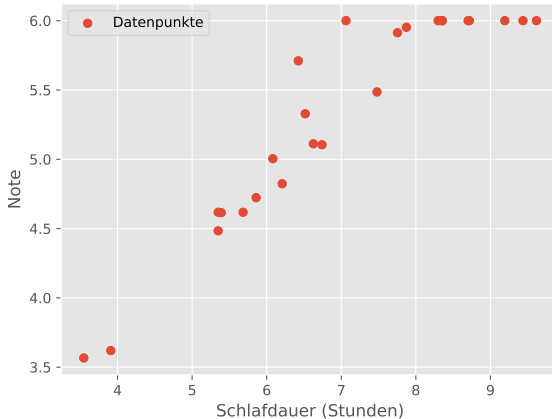
Variable: Note



FreeFlower

Zahlen vorhersagen mit Regression

Unabhängige und abhängige Variablen



► **Unabhängige Variable:**

Schlafdauer

► **Abhängige Variable:** Note

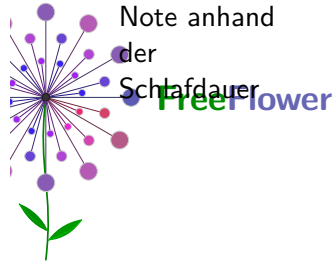
► **Ziel:** Vorhersage der Note anhand

der

Schlafdauer

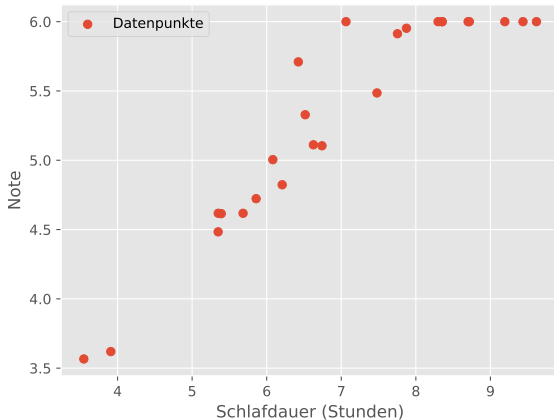
der

Schlafdauer



Zahlen vorhersagen mit Regression

Unabhängige und abhängige Variablen



► **Unabhängige**

Variable:

Schlafdauer

► **Abhängige**

Variable: Note

► **Ziel:**

Vorhersage der

Note anhand

der

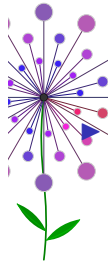
Schlafdauer

FreeFlower

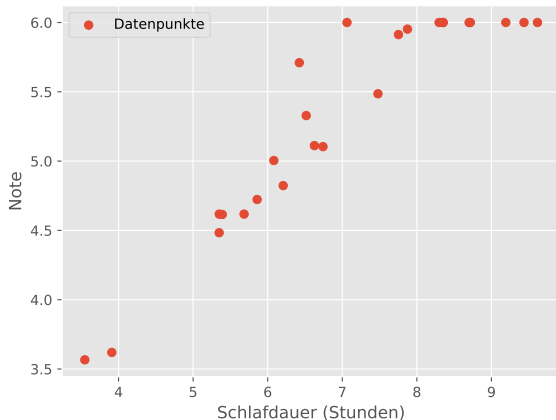
► **Beispiel:** 7.9

Stunden Schlaf

→ Note 5.5?



Lineare Funktion



Einfache Idee: zwei Datenpunkte verbinden, um eine Gerade zu berechnen:

Schlafdauer (=x)	Note (=y)
------------------	-----------

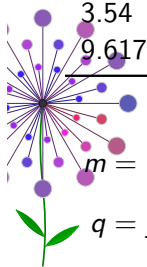
3.54	3.57
------	------

9.617	6.0
-------	-----

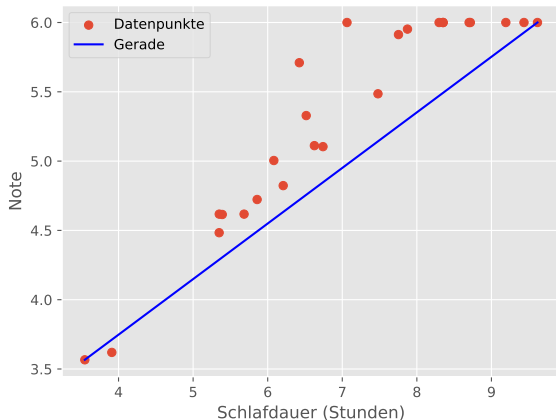
FreeFlower

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

$$q = y_1 - m \cdot x_1$$



Lineare Funktion



Einfache Idee: zwei Datenpunkte verbinden, um eine Gerade zu berechnen:

Schlafdauer (=x)	Note (=y)
------------------	-----------

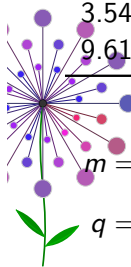
3.54	3.57
------	------

9.617	6.0
-------	-----

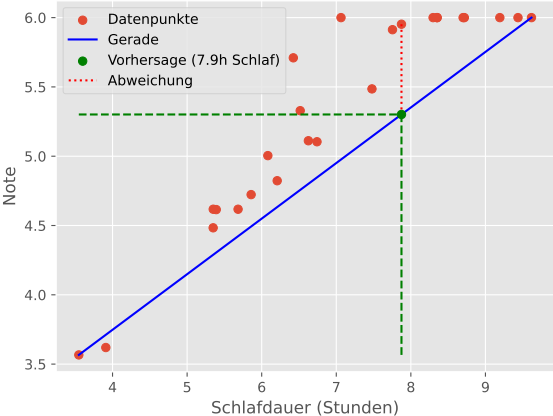
FreeFlower

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

$$q = y_1 - m \cdot x_1$$



Qualität der Gerade



	P_1	P_2	P_3
Schlaf	3.5	7.9	9.6
Note	3.6	6.0	6.0
Vorh.	3.6	5.3	6.0
Diff.	0.0	-0.7	0.0

Fehler:

$$0^2 + (-0.7)^2 + 0^2 = \underline{0.5}$$

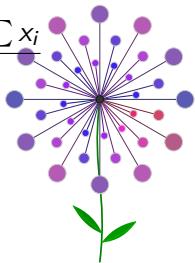
FreeFlower

Lineare Regression mit **OLS** - Ordinary Least Squares (Kleinste Quadrate)

$$m = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2}$$

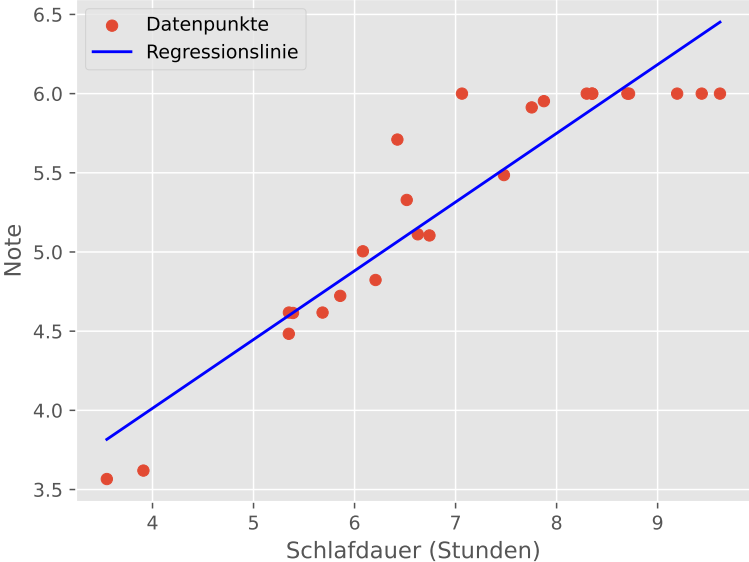
Das berechnete m ist die Steigung der Geraden, die den Fehler minimiert. Diese kann nun in die Gleichung für q eingesetzt werden:

$$q = \frac{\sum y_i - m \sum x_i}{n}$$



FreeFlower

OLS-Regressionslinie

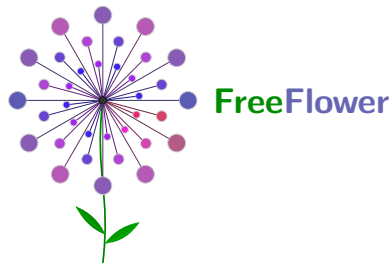


ower

Multiple lineare Regression

Manchmal hängt die Vorhersage nicht nur von einer Variable ab...

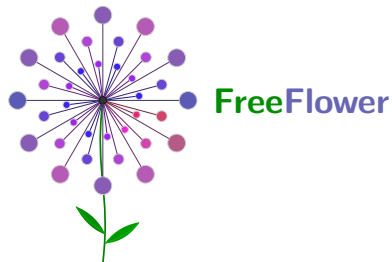
- ▶ Erfolgreicher Opernsänger 🎤



Multiple lineare Regression

Manchmal hängt die Vorhersage nicht nur von einer Variable ab...

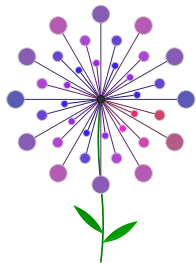
- ▶ Erfolgreicher Opernsänger 🎤
 - ▶ Schönes Aussehen 🕶️



Multiple lineare Regression

Manchmal hängt die Vorhersage nicht nur von einer Variable ab...

- ▶ Erfolgreicher Opernsänger 🎤
 - ▶ Schönes Aussehen 🕶️
 - ▶ Gute Stimme 🎵

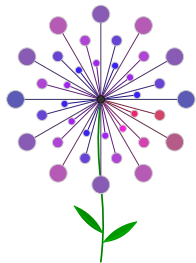


FreeFlower

Multiple lineare Regression

Manchmal hängt die Vorhersage nicht nur von einer Variable ab...

- ▶ Erfolgreicher Opernsänger 🎤
 - ▶ Schönes Aussehen 🕶️
 - ▶ Gute Stimme 🎵
- ▶ Autokauf 🚗

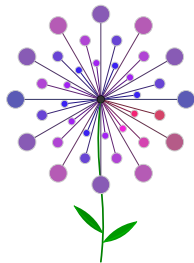


FreeFlower

Multiple lineare Regression

Manchmal hängt die Vorhersage nicht nur von einer Variable ab...

- ▶ Erfolgreicher Opernsänger 🎤
 - ▶ Schönes Aussehen 🕶️
 - ▶ Gute Stimme 🎵
- ▶ Autokauf 🚗
 - ▶ Preis 💰

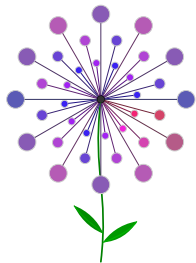


FreeFlower

Multiple lineare Regression

Manchmal hängt die Vorhersage nicht nur von einer Variable ab...

- ▶ Erfolgreicher Opernsänger 🎤
 - ▶ Schönes Aussehen 🕶️
 - ▶ Gute Stimme 🎵
- ▶ Autokauf 🚗
 - ▶ Preis 💰
 - ▶ Marke 🚗

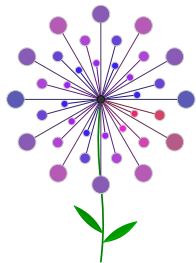


FreeFlower

Multiple lineare Regression

Manchmal hängt die Vorhersage nicht nur von einer Variable ab...

- ▶ Erfolgreicher Opernsänger 🎤
 - ▶ Schönes Aussehen 🕶️
 - ▶ Gute Stimme 🎵
- ▶ Autokauf 🚗
 - ▶ Preis 💰
 - ▶ Marke 🚗
 - ▶ Baujahr 📅 17



FreeFlower

Multiple lineare Regression

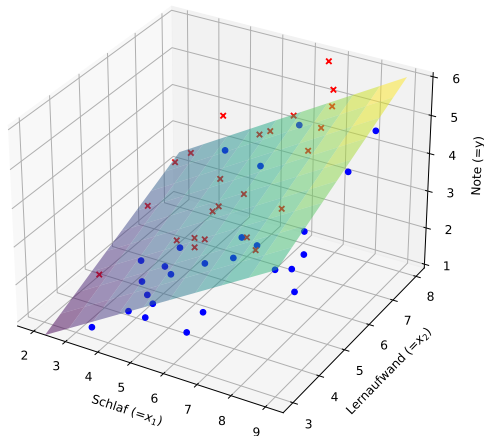
Mehrere unabhängige Variablen

- ▶ Beispiel: Note über Lernaufwand und Schlaf
- ▶ Erklärung:
 - ▶ x_1 : Lernaufwand
 - ▶ x_2 : Schlaf
 - ▶ y : Note

Formeln

- ▶ Einfache lineare Regression:

$$\begin{aligned}y &= q + mx \\ &= a_0 + a_1 x_1\end{aligned}$$



Multiple lineare Regression

Mehrere unabhängige Variablen

- ▶ Beispiel: Note über Lernaufwand und Schlaf
- ▶ Erklärung:
 - ▶ x_1 : Lernaufwand
 - ▶ x_2 : Schlaf
 - ▶ y : Note

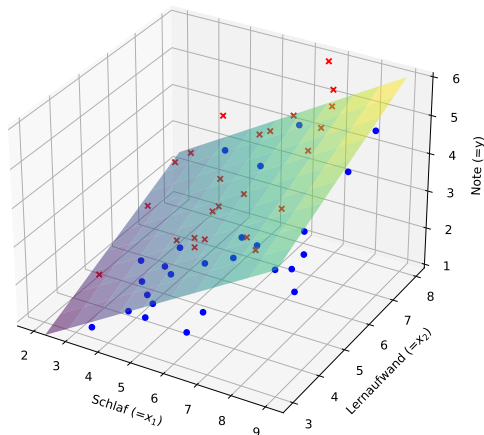
Formeln

- ▶ Einfache lineare Regression:

$$y = q + mx$$
$$= a_0 + a_1 x_1$$

- ▶ Multiple lineare Regression:

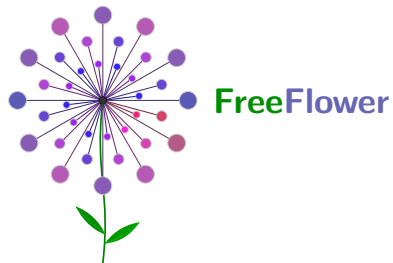
$$y = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3$$
$$+ \dots + a_n x_n$$



Auftrag

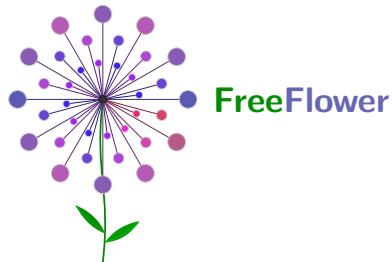
Jupyter Notebooks

- ▶ Aufgabenblatt 3a durchlesen und vervollständigen
- ▶ Aufgabenblatt 3b lösen



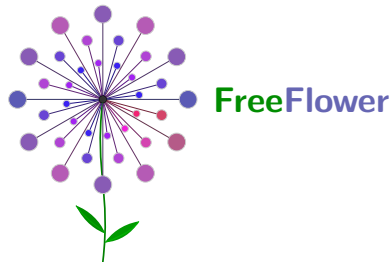
Lernziele

- ☑ Ich kann erklären, was eine **lineare Regression** ist und wofür sie verwendet wird.



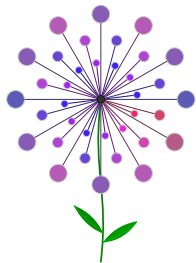
Lernziele

- ✓ Ich kann erklären, was eine **lineare Regression** ist und wofür sie verwendet wird.
- ✓ Ich kann den Unterschied zwischen **unabhängiger** und **abhängiger** Variable erläutern.



Lernziele

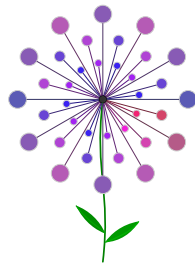
- ✓ Ich kann erklären, was eine **lineare Regression** ist und wofür sie verwendet wird.
- ✓ Ich kann den Unterschied zwischen **unabhängiger** und **abhängiger** Variable erläutern.
- ✓ Ich kann eine **lineare Funktion** aus zwei Punkten bestimmen.



FreeFlower

Lernziele

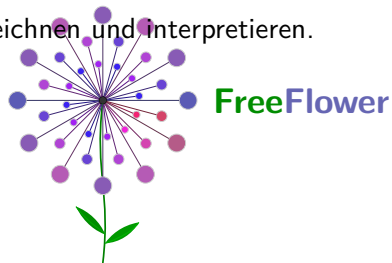
- ✓ Ich kann erklären, was eine **lineare Regression** ist und wofür sie verwendet wird.
- ✓ Ich kann den Unterschied zwischen **unabhängiger** und **abhängiger** Variable erläutern.
- ✓ Ich kann eine **lineare Funktion** aus zwei Punkten bestimmen.
- ✓ Ich kann das Prinzip der **Kleinste-Quadrate-Methode (OLS)** erklären und anwenden.



FreeFlower

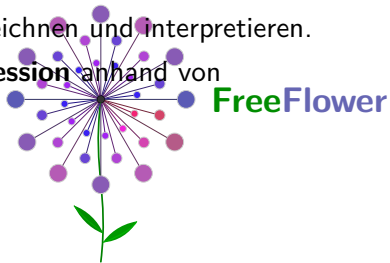
Lernziele

- ✓ Ich kann erklären, was eine **lineare Regression** ist und wofür sie verwendet wird.
- ✓ Ich kann den Unterschied zwischen **unabhängiger** und **abhängiger** Variable erläutern.
- ✓ Ich kann eine **lineare Funktion** aus zwei Punkten bestimmen.
- ✓ Ich kann das Prinzip der **Kleinste-Quadrate-Methode (OLS)** erklären und anwenden.
- ✓ Ich kann eine Regressionsgerade zeichnen und interpretieren.



Lernziele

- ✓ Ich kann erklären, was eine **lineare Regression** ist und wofür sie verwendet wird.
- ✓ Ich kann den Unterschied zwischen **unabhängiger** und **abhängiger** Variable erläutern.
- ✓ Ich kann eine **lineare Funktion** aus zwei Punkten bestimmen.
- ✓ Ich kann das Prinzip der **Kleinste-Quadrate-Methode (OLS)** erklären und anwenden.
- ✓ Ich kann eine Regressionsgerade zeichnen und interpretieren.
- ✓ Ich kann die **Qualität einer Regression** anhand von Fehlermassen beurteilen.



Lernziele

- ✓ Ich kann erklären, was eine **lineare Regression** ist und wofür sie verwendet wird.
- ✓ Ich kann den Unterschied zwischen **unabhängiger** und **abhängiger** Variable erläutern.
- ✓ Ich kann eine **lineare Funktion** aus zwei Punkten bestimmen.
- ✓ Ich kann das Prinzip der **Kleinste-Quadrate-Methode (OLS)** erklären und anwenden.
- ✓ Ich kann eine Regressionsgerade zeichnen und interpretieren.
- ✓ Ich kann die **Qualität einer Regression** anhand von Fehlermassen beurteilen.
- ✓ Ich kann eine lineare Regression in **Python** durchführen.

