

The background of the slide features several large, overlapping, irregular blue shapes that resemble stylized stones or pebbles. These shapes are positioned primarily on the left and bottom sides, leaving the top right area clear for the text.

# OLS im Vergleich zu Lasso und Ridge Regression

Vorlesung Dr. Andreas Gocht-Zech  
<[andreas.gocht-zech@posteo.de](mailto:andreas.gocht-zech@posteo.de)>

# Inhalt

- Einführung
- Ordinary Least Squares (OLS)
- Ridge-Regression
- Lasso-Regression
- Zusammenfassung

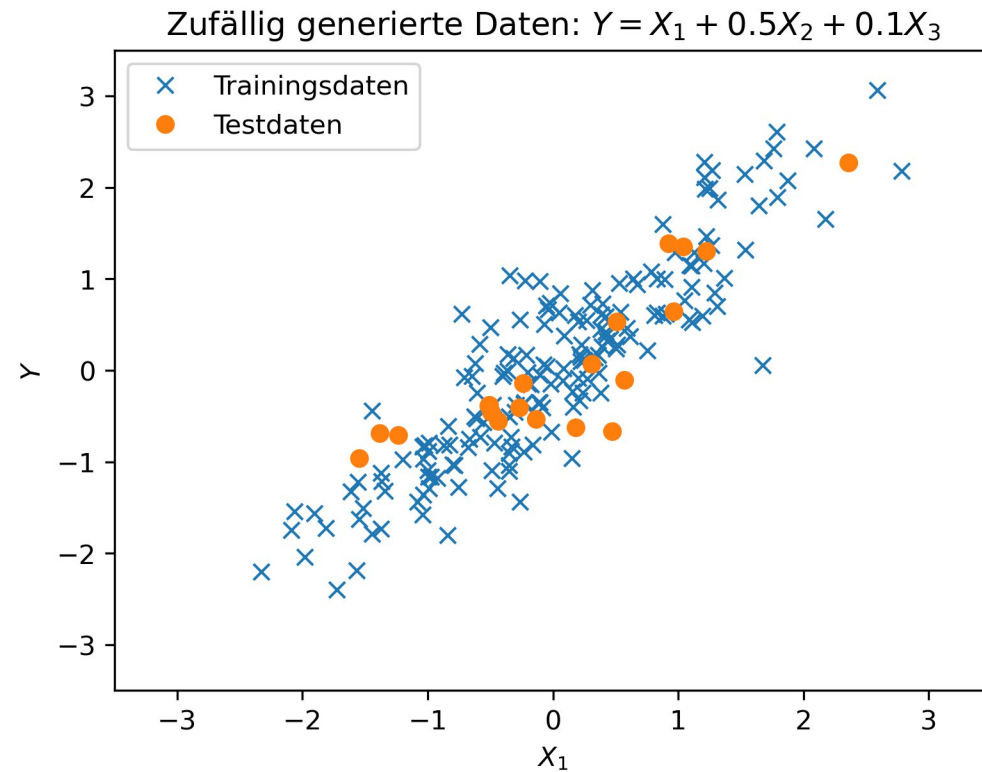
- Regression:
  - Standardmethode der Statistik:  
Frage nach der Beziehung zwischen einer abhängigen Variable  $y$  und einer oder mehreren unabhängiger Variable  $(X = \{x_1, x_2, \dots, x_n\})$

- Genutzt für:
  - Modellierung:
    - Bestimmung von Parametern basierend auf Messwerten
    - Neuronale Netze
  - Analysen
    - Identifikation von Faktoren, die ein Ergebnis beeinflussen z.B. bei Studien in der Medizin und der Psychologie

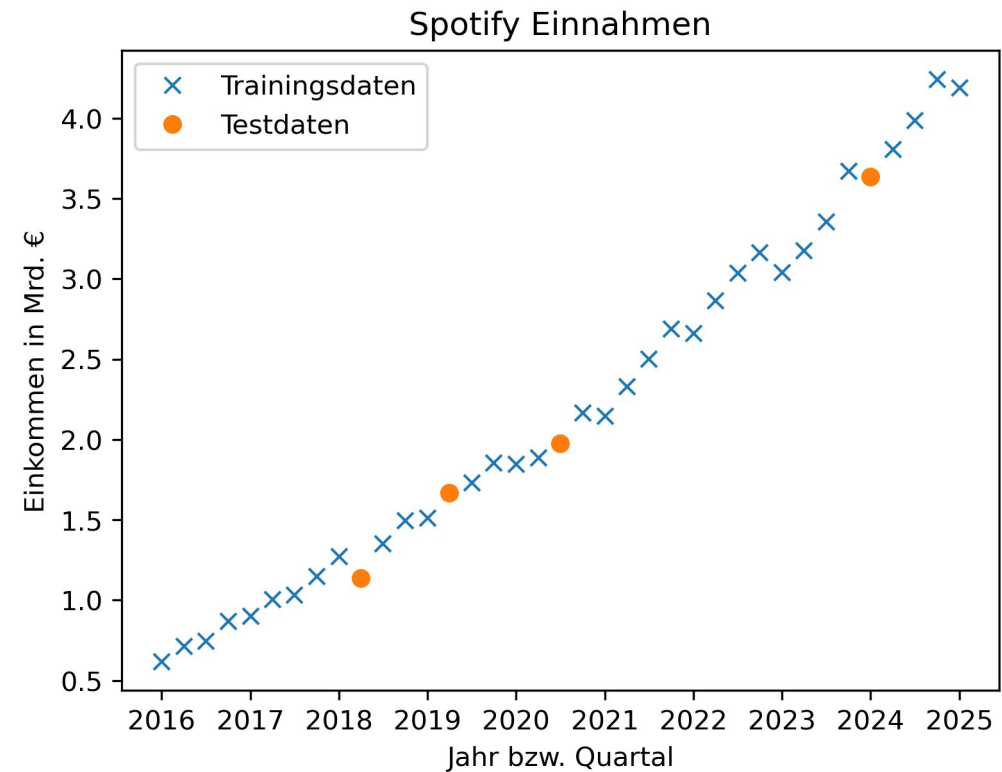
- Synthetisch generierte Daten:
  - $y = x_1 + 0.5x_2 + 0.1x_3$
  - $x_{1,2,3} \in N(0,1)$
- Spotify Daten:
  - Einnahmen ( $y$ )
  - Monatlich Aktive Nutzer
  - Premium Nutzer
  - In den USA steuerpflichtige Einnahmen aus den Bereichen Werbung, PR, usw.

# Beispiel

## Synthetisch



## Spotify



# Ordinary Least Squares (OLS) Eindimensional

- Einfache Regression:

$$\hat{y} = \hat{f}(x) = b \cdot x_1 + a$$

Wobei  $b$  und  $a$  zu bestimmen sind

- Ziel: Minimierung der Fehlerfunktion:

$$E = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

# Ordinary Least Squares (OLS) Eindimensional

Wie Minimiere ich eine Fehlerfunktion?



# Ordinary Least Squares (OLS) Eindimensional

- Minimierung:  $\frac{dE}{db} = 0; \frac{dE}{da} = 0$

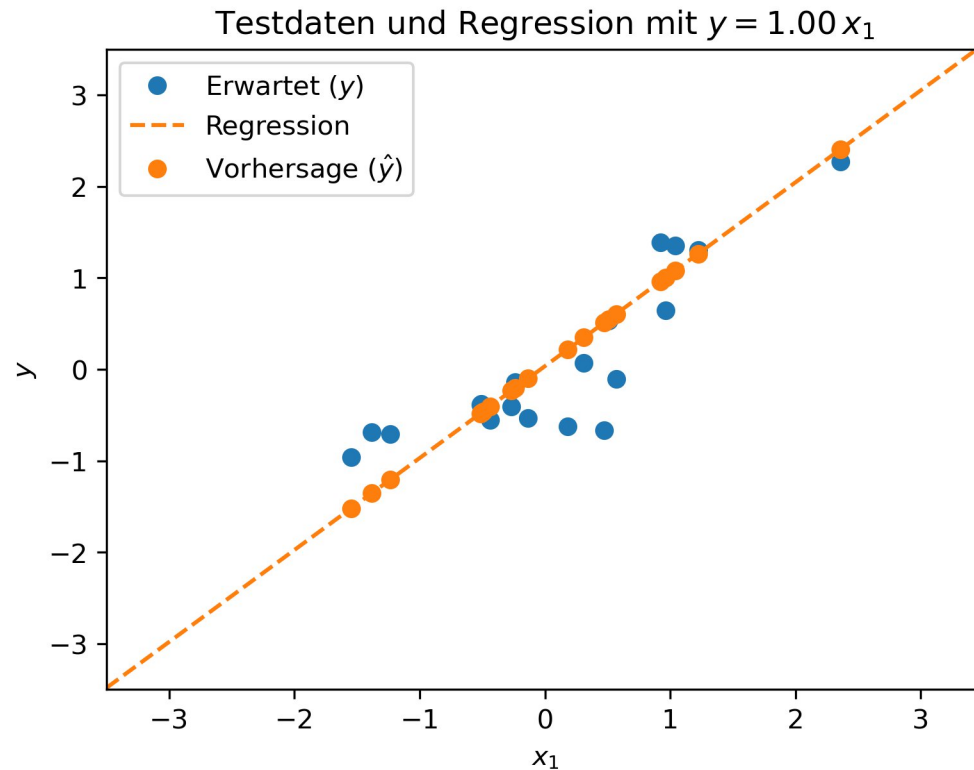
- Lösung:  $a = \bar{y} - b \cdot \bar{x}$

$$b = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{S_{xy}}{\sigma_x}$$

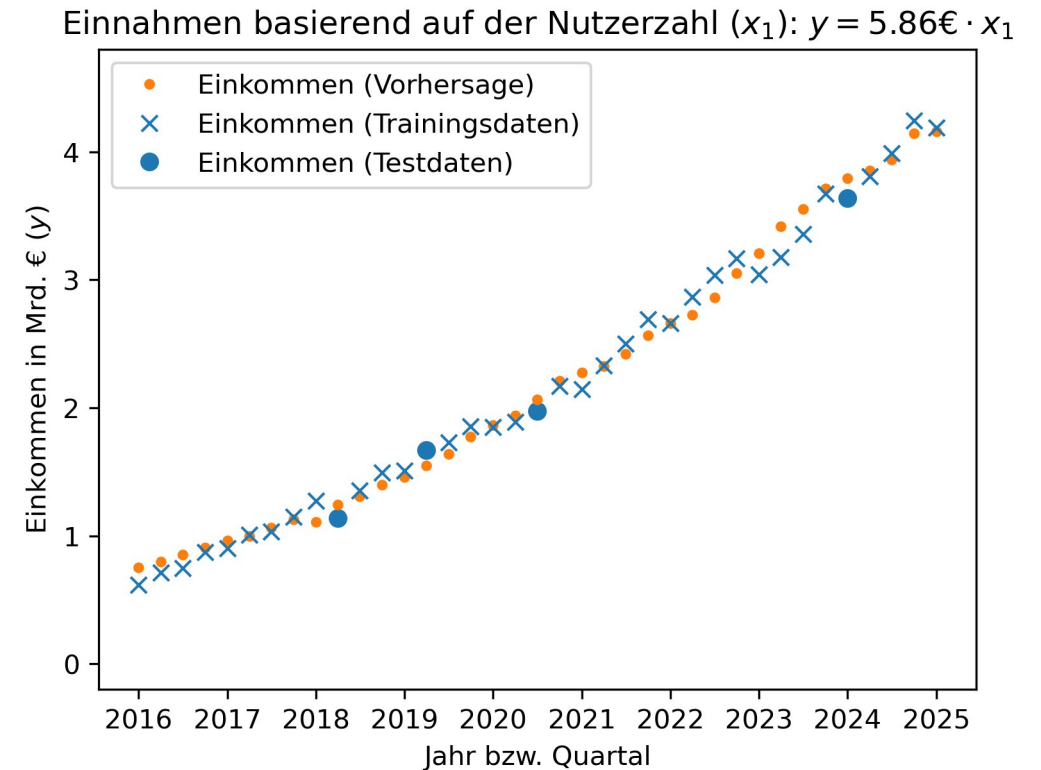
- Wobei:  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i ; \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$

# Ordinary Least Squares (OLS) Beispiel

## Synthetisch



## Spotify



Three overlapping, irregular gray shapes in the top-left corner of the slide.

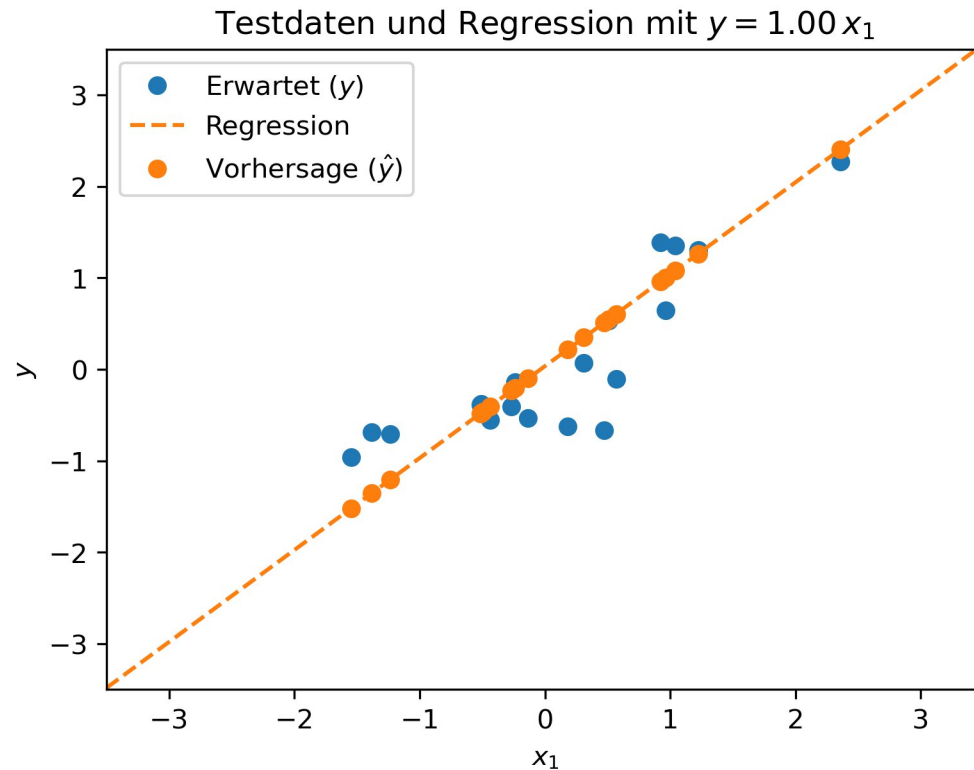
# Ordinary Least Squares (OLS) Eindimensional

Warum ist die Spotify-Regression keine gerade?

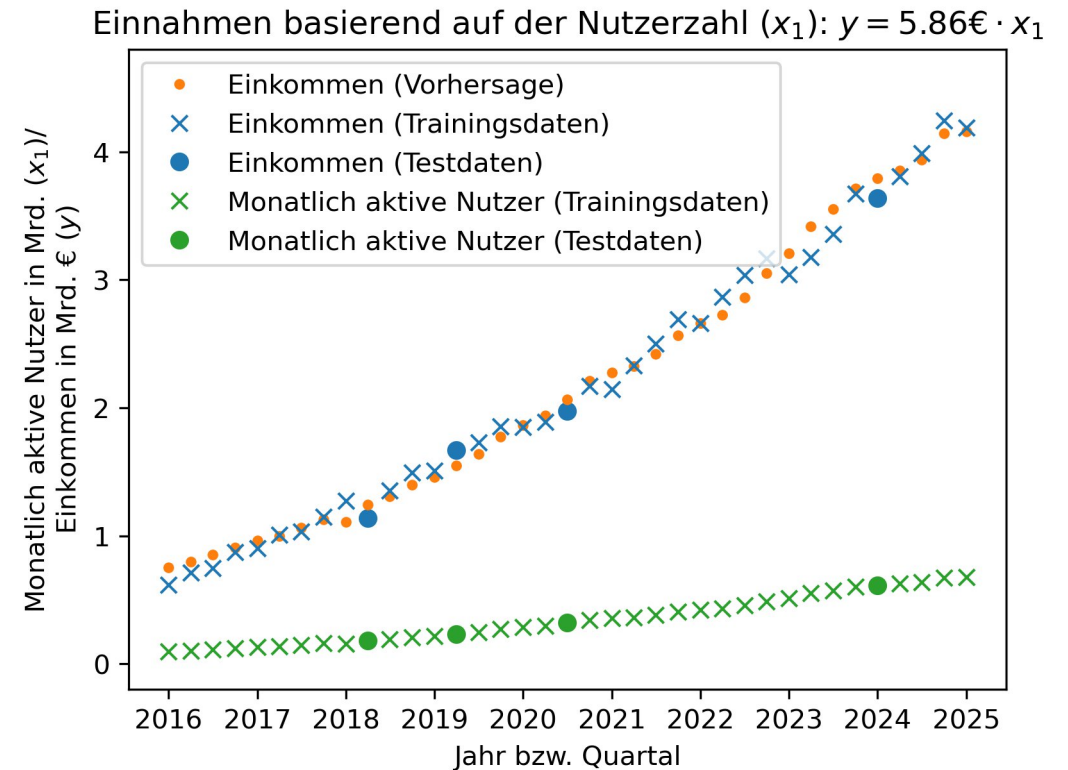
# Ordinary Least Squares (OLS)

## Beispiel

### Synthetisch



### Spotify



# Ordinary Least Squares (OLS) Mehrdimensional

- Angepasste Regressionsfunktion:

$$\hat{y} = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_m \cdot x_m$$

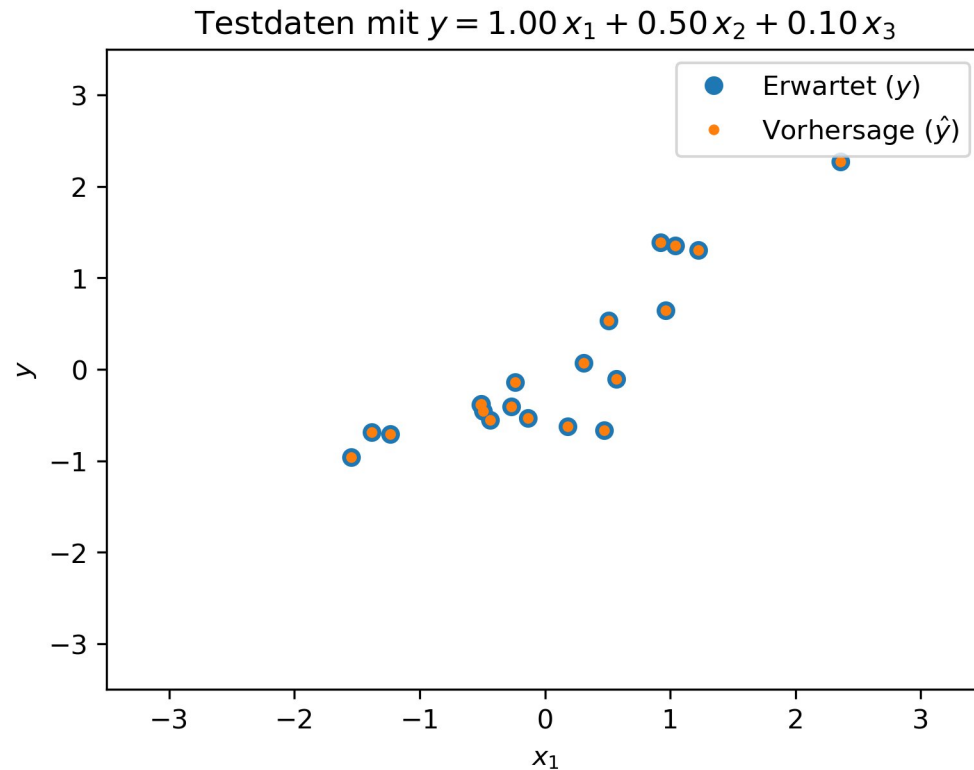
- Fehlerfunktion bleibt gleich:

$$E = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

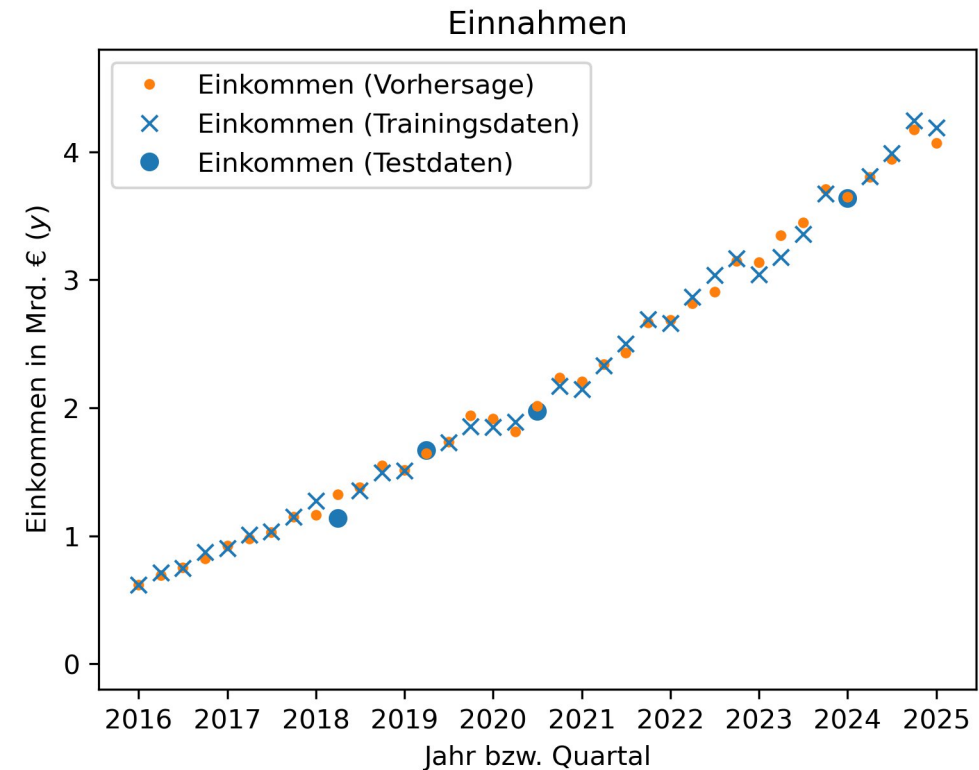
- Lösung über numerische Verfahren (z.B. Gradientenverfahren)

# Ordinary Least Squares (OLS) Beispiel

## Synthetisch



## Spotify



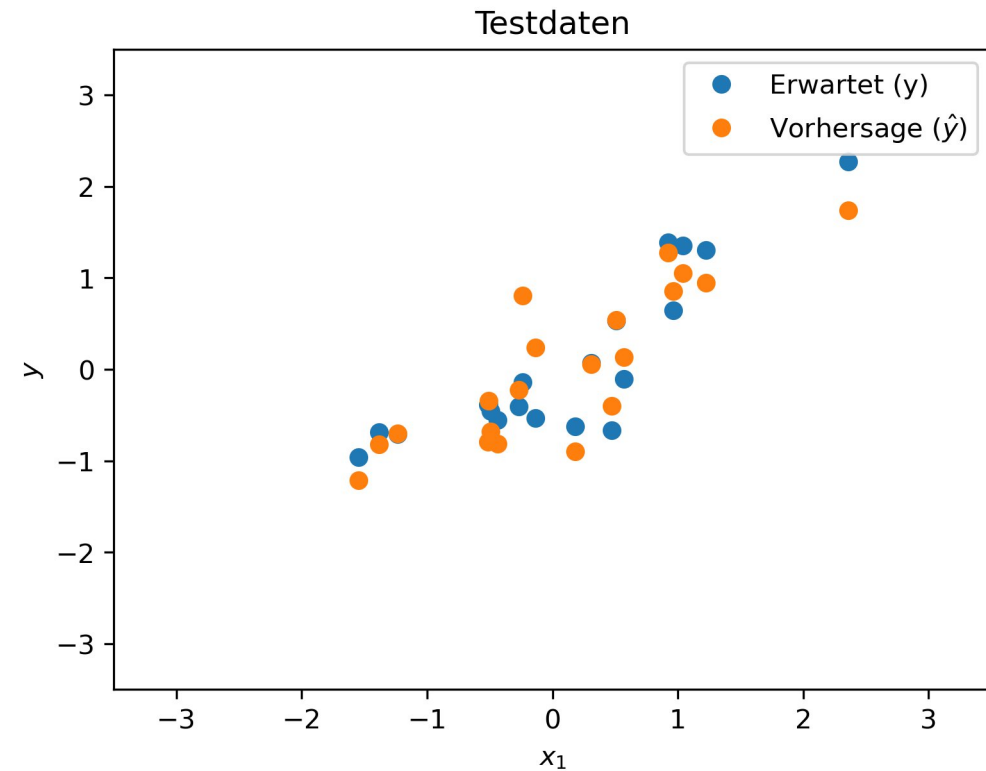
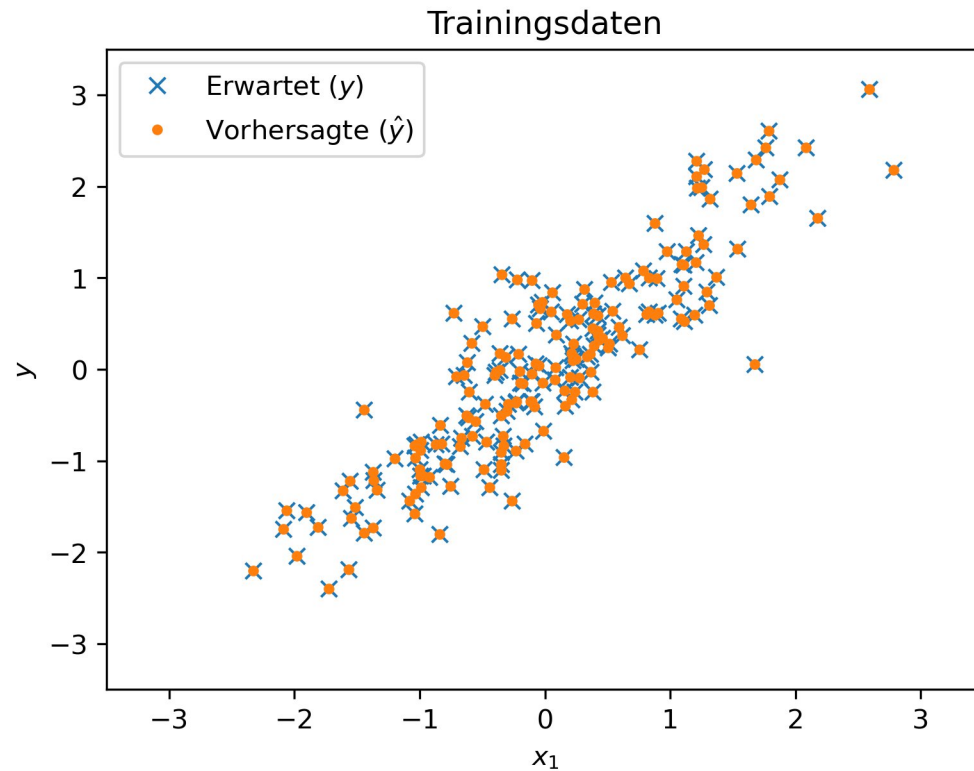
# Ridge-Regression

## Overfitting

- Bei vielen Parametern  $b$  kann es zu einer Überanpassung kommen (Overfitting)
  - Der Fehler auf den Trainingsdaten ist gering
  - Der Fehler auf den Testdaten ist hoch
- Beispiel:
  - Wir fügen zu unserer synthetischen Funktion 200 zufällig generierte unabhängige hinzu:

$$x_{4,\dots,204} \in N(0,1)$$

# Ridge-Regression Overfitting





# Ridge-Regression Overfitting

- Spotify Daten auf den Mittelwert normieren

$$y_{i,\text{norm}} = \frac{y_i}{\bar{y}}; x_{i,\text{norm}} = \frac{x_i}{\bar{x}_i}$$

- $x_1$ ... Einnahmen basierend auf der monatlich aktiven Nutzer
- $x_2$ ... Premium Nutzer
- $x_3$ ... Quartal
- $x_4$ ... Jahr
- $x_5$ ... Einnahmen aus Werbung in den USA
- $y = -242.31 + 0.41x_1 - 0.35x_2 + 0.05x_3 + 242.63x_4 + 0.58x_5$

# Ridge-Regression Overfitting

- Welche Terme sind vermutlich nicht Sinnvoll?
  - $x_1$ ... Einnahmen basierend auf der monatlich aktiven Nutzer
  - $x_2$ ... Premium Nutzer
  - $x_3$ ... Quartal
  - $x_4$ ... Jahr
  - $x_5$ ... Einnahmen aus Werbung in den USA
- $y = -242.31 + 0.41x_1 - 0.35x_2 + 0.05x_3 + 242.63x_4 + 0.58x_5$

# Ridge-Regression

- Lösung: Regularisierung
  - „Strafterm“ für viele Parameter
  - Ridge nutzt die sogenannte L2-Norm

$$E_{\text{Ridge}} = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \alpha \cdot \sum_{j=1}^m b_j^2$$

# Ridge-Regression Beispiel

- Lineare Regression:

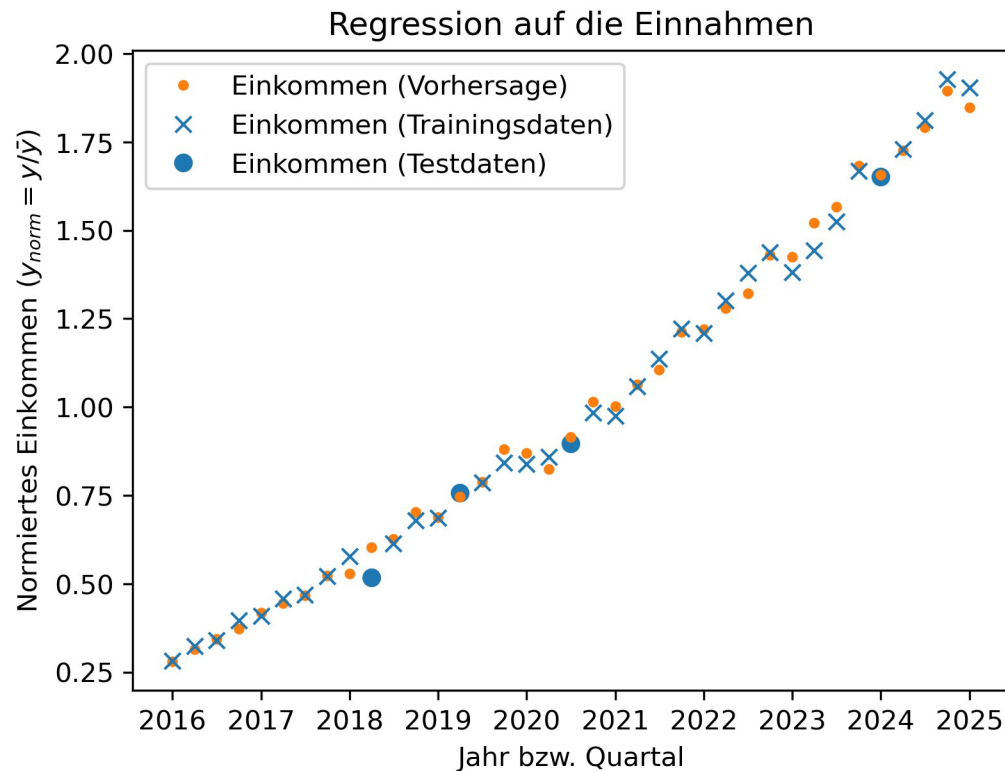
$$y = -242.31 + 0.41x_1 - 0.35x_2 + 0.05x_3 + 242.63x_4 + 0.58x_5$$

- Ridge Regression ( $\alpha = 1$ ):

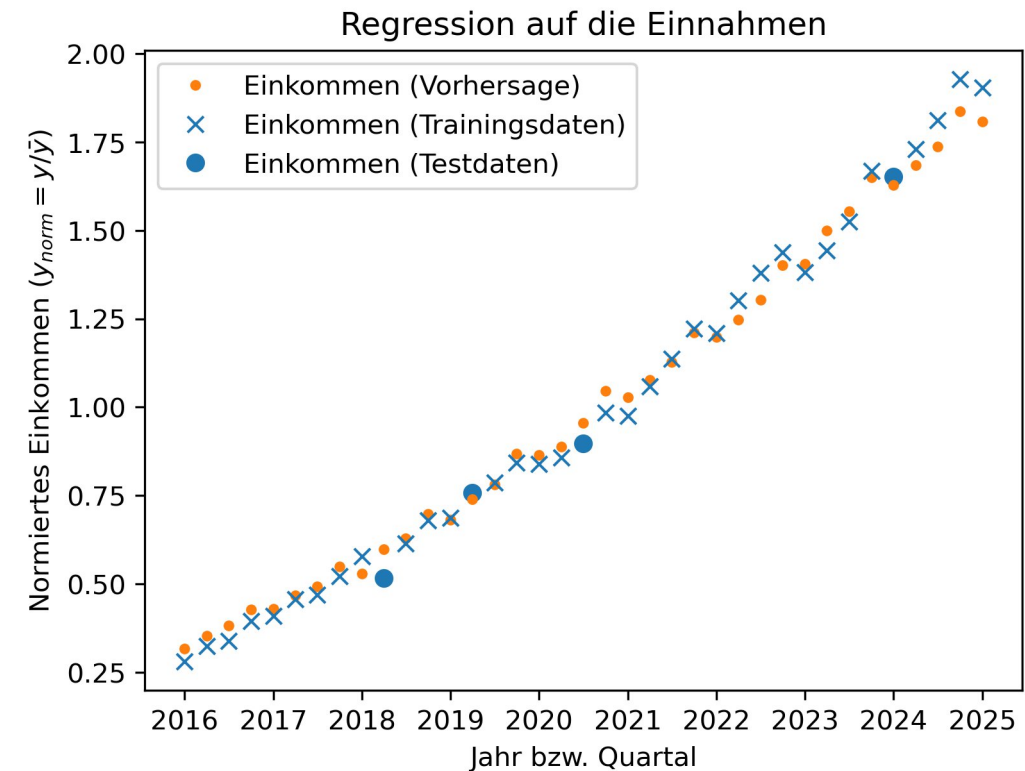
$$y = -0.02 + 0.43x_1 - 0.39x_2 + 0.01x_3 + 0.00x_4 + 0.19x_5$$

# Ridge-Regression Beispiel

## Lineare-Regression

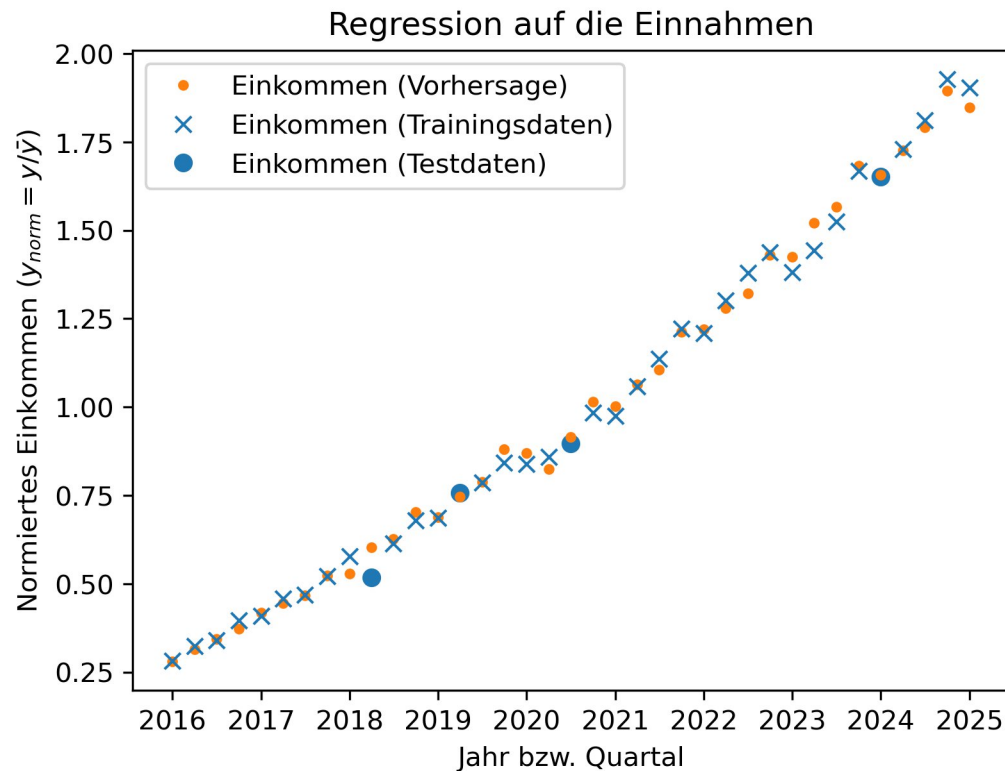


## Ridge-Regression

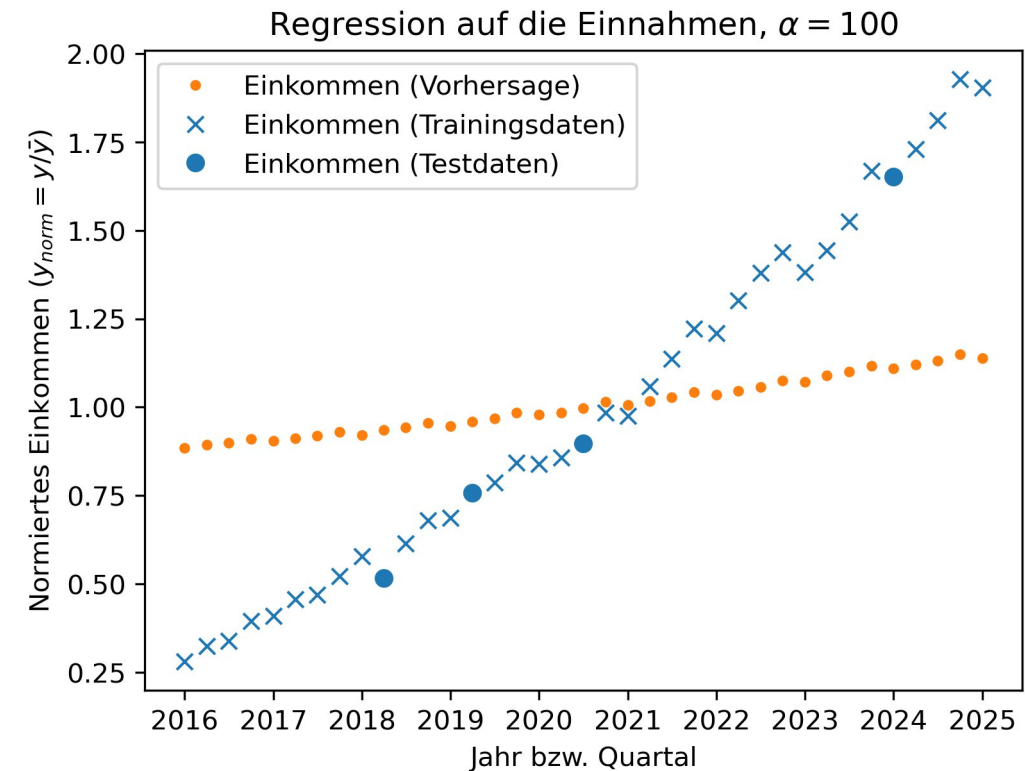


# Ridge-Regression Beispiel

## Lineare-Regression



## Ridge-Regression



# Lasso-Regression

- Ähnlich zu Ridge, nur mit L1-Norm

$$E_{\text{Lasso}} = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \alpha \cdot \sum_{j=1}^m |b_j|$$

- Tendiert zum „Nullsetzen“ der Terme

# Lasso-Regression Beispiel

- Lineare Regression:

$$y = -242.31 + 0.41x_1 - 0.35x_2 + 0.05x_3 + 242.63x_4 + 0.58x_5$$

- Ridge-Regression ( $\alpha = 1$ ):

$$y = -0.02 + 0.43x_1 - 0.39x_2 + 0.01x_3 + 0.00x_4 + 0.19x_5$$

- Lasso-Regression ( $\alpha = 1$ ):

$$y = 0.13 + 0.79x_1 - 0.08x_2 + 0.00x_3 + 0.00x_4 + 0.00x_5$$

Zur Erinnerung:

$x_1$ ... Einnahmen basierend auf der monatlich  
aktiven Nutzer

$x_2$ ... Premium Nutzer

$x_3$ ... Quartal

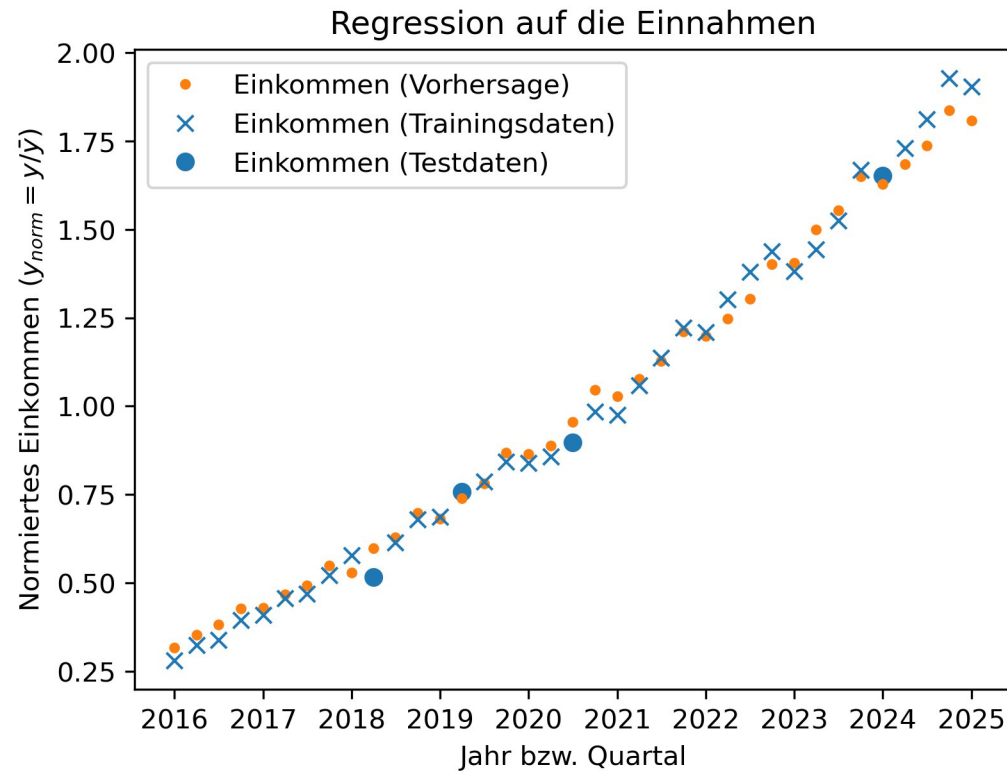
$x_4$ ... Jahr

$x_5$ ... Einnahmen aus Werbung in den USA

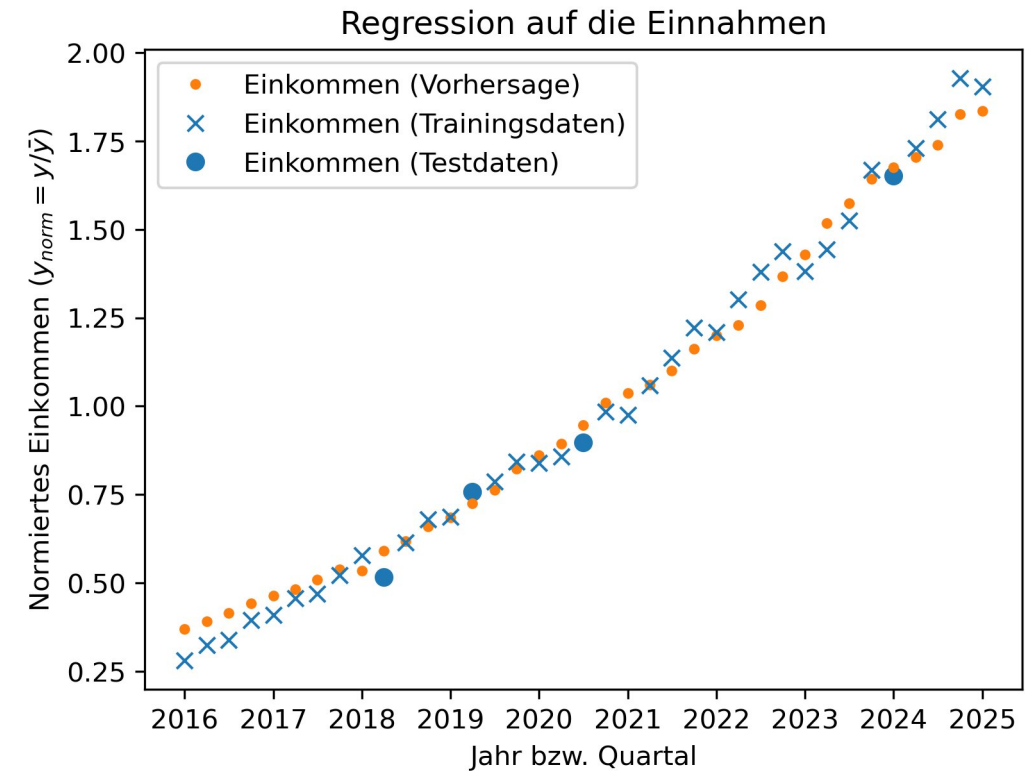


# Lasso-Regression Beispiel

## Ridge-Regression – Spotify



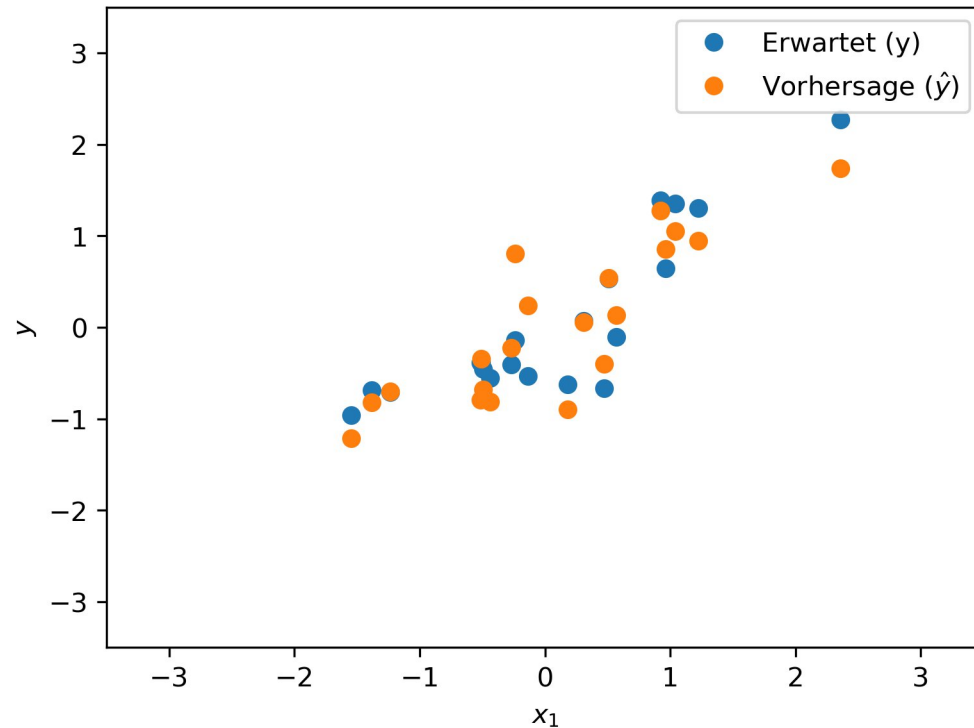
## Lasso-Regression – Spotify



# Lasso-Regression Beispiel

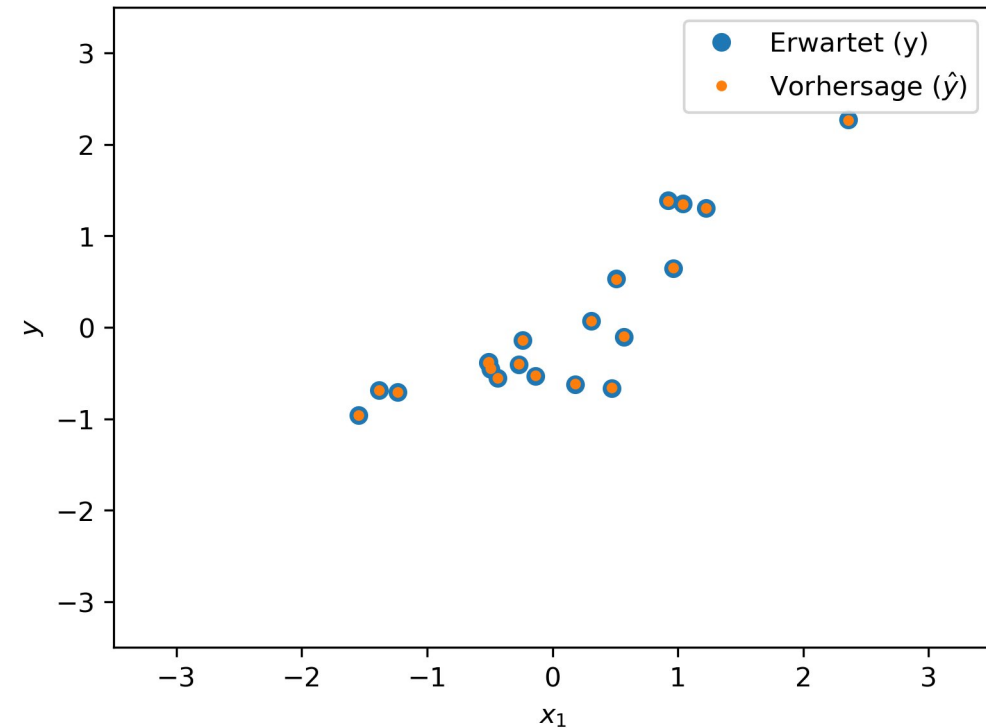
## Lineare Regression – Synthetisch

Vorhersage auf Testdaten mit  $\hat{y} = 0.87x_1 + 0.44x_2 + 0.14x_3$   
und 200 Zufällig generierten Vektoren



## Lasso Regression – Synthetisch

Vorhersage auf Testdaten mit  $\hat{y} = 1.00x_1 + 0.50x_2 + 0.10x_3$   
und 200 Zufällig generierten Vektoren



# Zusammenfassung

- Einführung der Linearen Regression
- Overfitting
- Regularisierung mit
  - Ridge
  - Lasso

- Ridge und Lasso-Regression lassen sich kombinieren:

$$E_{\text{Elastic-Net}} = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \alpha_1 \cdot \sum_{j=1}^m |b_j| + \alpha_2 \cdot \sum_{j=1}^m b_j^2$$

# Ausblick

## Mehrdimensional und Polynomial

- Angepasste Regressionsfunktion:

$$\begin{aligned}\hat{y} = b_0 &+ b_{1,1} \cdot x_1 + b_{2,1} \cdot x_2 + \dots + b_{m,1} \cdot x_m \\ &+ b_{1,2} \cdot x_1^2 + b_{2,2} \cdot x_2^2 + \dots + b_{m,2} \cdot x_m^2 \\ &+ \dots \\ &+ b_{1,k} \cdot x_1^k + b_{2,k} \cdot x_2^k + \dots + b_{m,k} \cdot x_m^k\end{aligned}$$

- Anstelle der polynomialen Terme lassen sich auch beliebige andere Funktionen nutzen

# Vielen Dank

Python Quellcode  
der Beispiele:

- Fragen?
  - Gerne auch per Mail:  
[andreas.gocht-zech@posteo.de](mailto:andreas.gocht-zech@posteo.de)
  - Oder GitHub Ticket



[https://github.com/AndreasGocht/  
Lineare-Regression/](https://github.com/AndreasGocht/Lineare-Regression/)