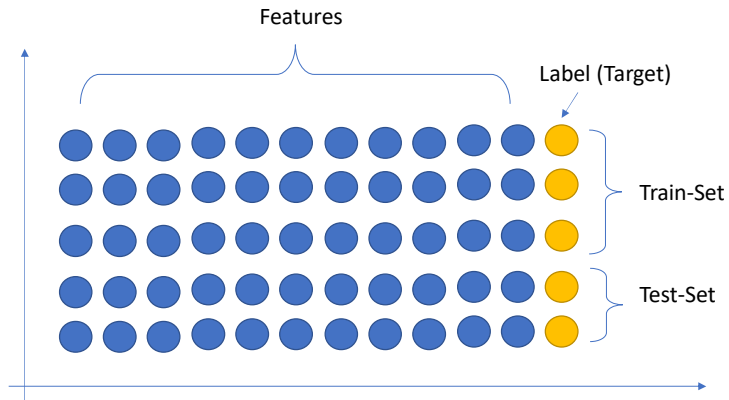


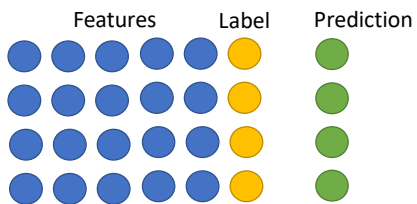
Regressionsanalyse

- Dataset: Train- und Test-Set eingeteilt
- Ein Modell wird erstellt und mit Train-Set trainiert



1

Modelle evaluieren/auswerten

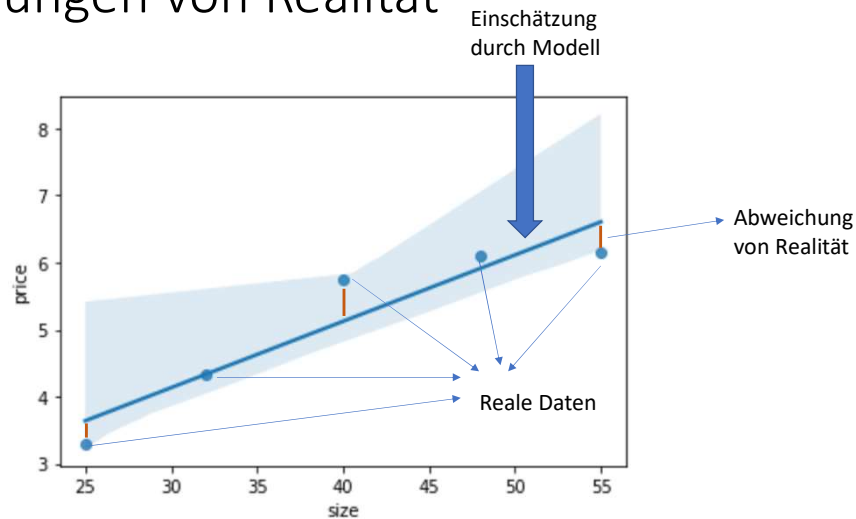


Um die Genauigkeit eines Modells auszuwerten bzw. das Modell zu evaluieren:

- Das Modell wird mit gewissen Datensätzen trainiert
- Dann lassen wir das Modell für diese Daten eine Einschätzung machen
- Dann werden diese Einschätzung mit realen Angaben verglichen, um zu sehen, inwiefern das Modell die Werte gut und richtig eingeschätzt hat.

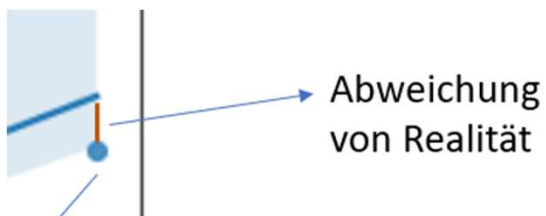
2

Abweichungen von Realität



3

Abweichung von Realität



	size	price	prediction
0	25	3.28	3.637311
1	32	4.33	4.330166
2	40	5.75	5.122000
3	48	6.10	5.913834
4	55	6.15	6.606689

Zwei Punkte sind an beiden Enden des roten Segmentes zu erkennen:

Der echte Datenpunkt in der Realität: (55, 6.15)

Die Einschätzung an der blauen Gerade: (55, 6.606689)

4

Metrics (Maßzahlen)

In der Statistik gibt es verschiedene Algorithmen, die diese Abweichung beschreiben, u. a.

- Mean Absolute Error (MAE)
- Mean Squared Errors (MSE)
- Root Mean Squared Error (RMSE)

Usw.

R-Squared ist dabei eine wichtige Maßzahl

$$MAE = \frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}_j|$$

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

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2}$$

$$RAE = \frac{\sum_{j=1}^n |y_j - \hat{y}_j|}{\sum_{j=1}^n |y_j - \bar{y}|}$$

$$RSE = \frac{\sum_{j=1}^n (y_j - \hat{y}_j)^2}{\sum_{j=1}^n (y_j - \bar{y})^2}$$

$$R^2 = 1 - RSE$$