

Regularisierung von linearer Regression

Phillip Grafendorfer, Michael Kastner, Raphael Peer

Daten

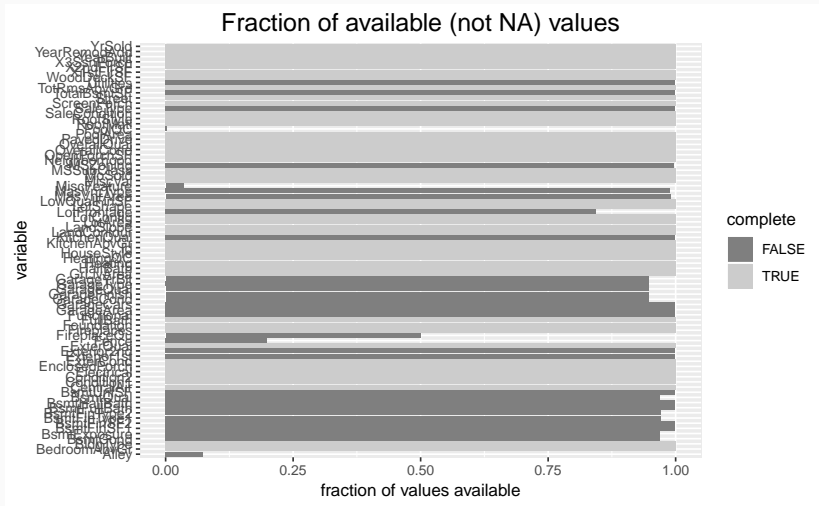
Ames House price Dataset

Datensatz:

- 1460 Häuser
- 79 erklärende Variablen (numerisch und kategorisch)
- bekannter Übungsdatensatz

Quelle: <https://www.kaggle.com/c/house-prices-advanced-regression-techniques>

Fehlende Werte: Übersicht



Fehlende Werte: Strategie

Umgang mit fehlenden Werten:

- Bei mehr als 10% Fehlenden Werten: Variable verworfen
- Bei numerischen Variablen: NA durch Median der Variable ersetzt
- Bei kategorischen Variablen: NA als eigene Kategorie (Kategorie 'unbekannt')

Problem mit validation-set: seltene Factor-levels

data-frame

einige levels im validation-set aber
nicht im trainings-set

⇒ unbekannte dummy

Variablen im validation-set

⇒ error

design-matrix

einige levels in
validation-set aber nicht im
trainings-set

⇒ dummy variable

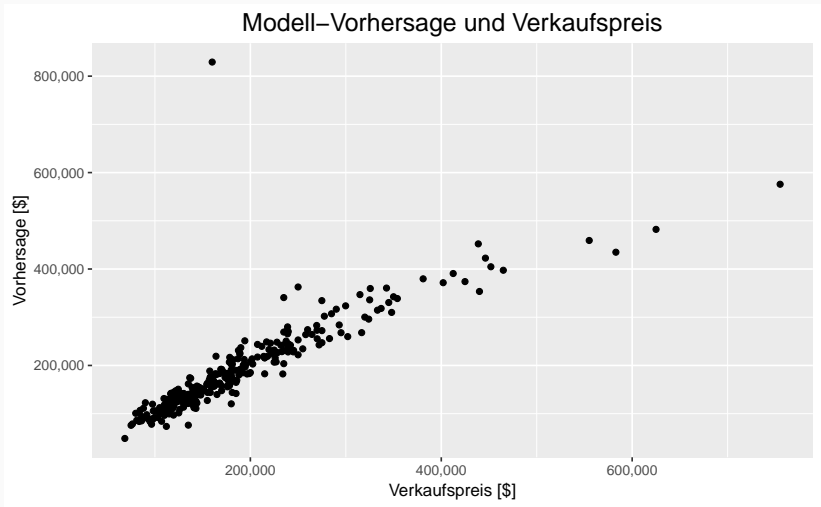
immer null im trainings set

⇒ Koeffizient ≈ 0

⇒ kein Einfluss

Standard lineare Regression

Einfaches Model mit allen Variablen



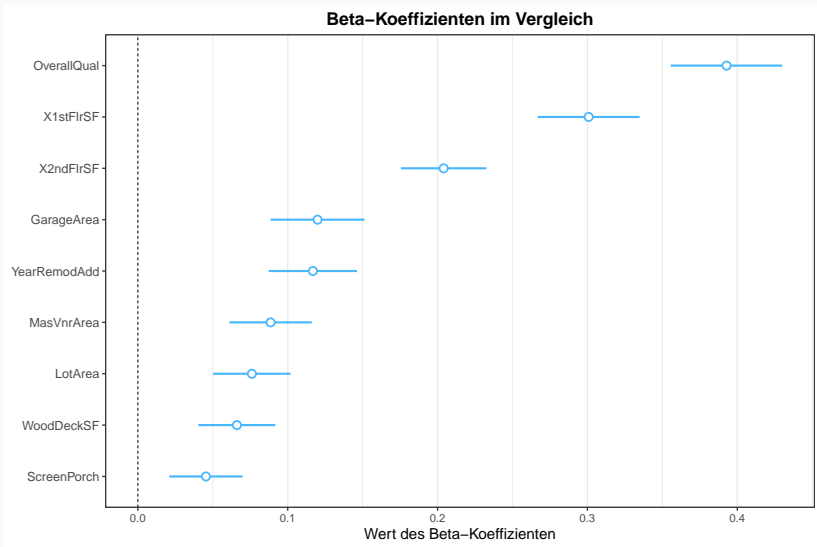
Interpretierbare Koeffizienten

Nachteil unstandardisierter Regressionskoeffizienten

- Von den Maßeinheiten für X und Y abhängig
- Daher schlechtere Vergleichbarkeit

Lösung: Standardisierte Koeffizienten

Beta-Koeffizienten im Vergleich



Regularisierung

Problemstellung I

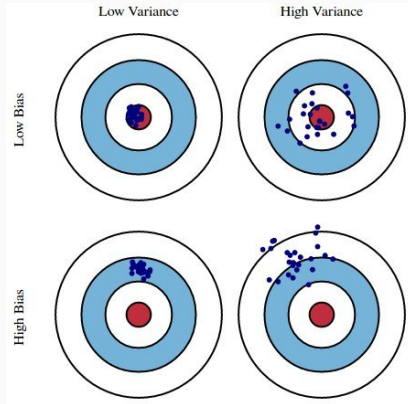


Figure 1: Quelle: kdnuggets.com

- Bias- Variance Tradeoff
- OLS Schätzer ist "unbiased" aber kann große Varianz haben

Problemstellung II

Wann tritt große Varianz auf?

- Wenn die Prediktoren hohe Korrelation aufweisen
- Bei vielen Prediktoren. Wenn die Anzahl Prediktoren nahe bei Anzahl der Beobachtungen geht die Varianz gegen unendlich.

Lösung

Verringerung der Varianz auf Kosten des Bias

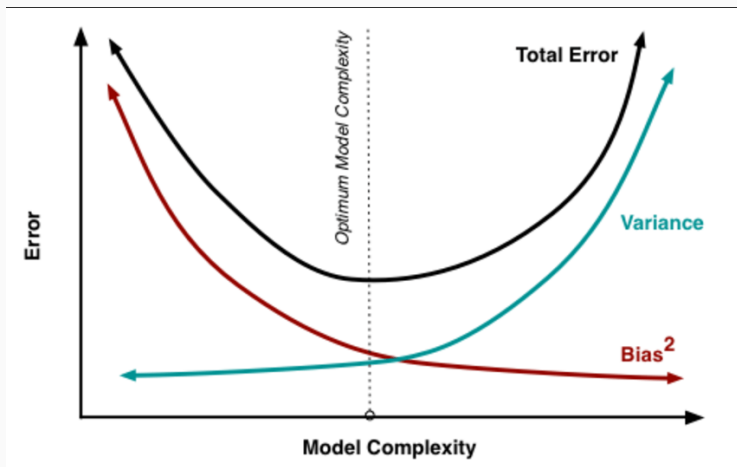


Figure 2: Quelle: researchgate.net

Ridge Regression

$$L_{ridge}(\hat{\beta}) = \sum_{i=1}^n (y_i - x_i' \hat{\beta})^2 + \lambda \sum_{j=1}^m \hat{\beta}_j^2 = \|y - X\hat{\beta}\|^2 + \lambda \| \hat{\beta} \|^2$$

Die Diskussion dieser Likelihood Funktion liefert für jeden Parameter λ ein set von Schätzern $\hat{\beta}$. Falls $\lambda \Rightarrow 0$, dann $\hat{\beta}_{ridge} \Rightarrow \hat{\beta}_{OLS}$ Frage: wie wird der Regularisierungs- Parameter gewählt?

- Crossvalidierung (hier benutzt)
- Minimierung eines weiteren Informationskriteriums (AIC, BIC etc.)

Ridge Regression: Crossvalidierung

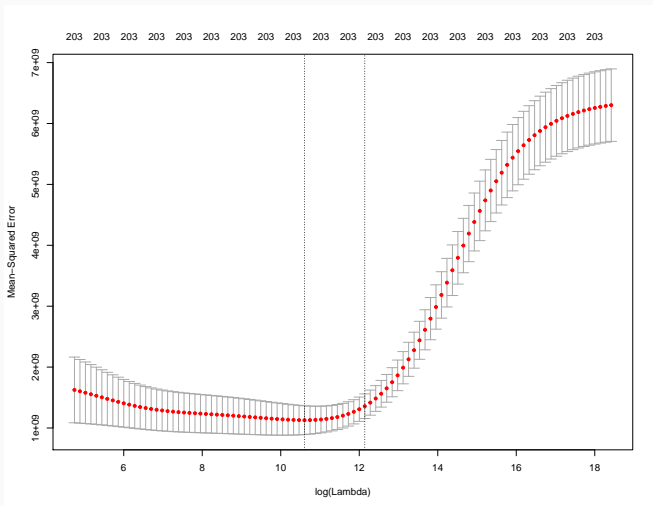


Figure 3: Lambda Tuning

Lasso Regression

$$L_{lasso}(\hat{\beta}) = \sum_{i=1}^n (y_i - x_i' \hat{\beta})^2 + \lambda \sum_{j=1}^m |\hat{\beta}_j| = \|y - X\hat{\beta}\|^2 + \lambda \|\hat{\beta}\|$$

Die Diskussion dieser Likelihood Funktion liefert für jeden Parameter λ ein set von Schätzern $\hat{\beta}$. Falls $\lambda \Rightarrow 0$, dann $\hat{\beta}_{lasso} \Rightarrow \hat{\beta}_{OLS}$ Frage: wie wird der Regularisierungs- Parameter gewählt?

- Crossvalidierung (hier benutzt)
- Minimierung eines weiteren Informationskriteriums (AIC, BIC etc.)

Lasso Regression: Crossvalidierung

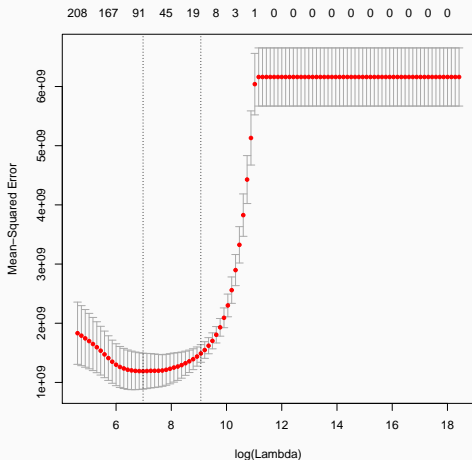


Figure 4: Lambda Tuning

Lasso Regression: Koeffizienten

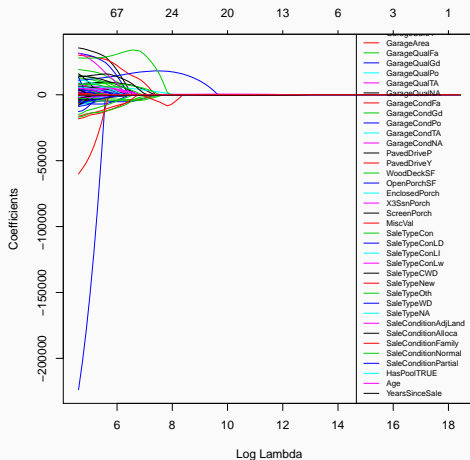


Figure 5: Lambda Tuning

Vergleich der Modelle

Modelle	R^2	MAD
OLS	0.933	20117
Ridge	0.903	19624
Lasso	0.904	30010

Vielen Dank für
eure Aufmerksamkeit!