

Topic 4: Regularisierung – Lasso und Ridge
Block 3: Komplexitaet, Kausalitaet & Generalisierung

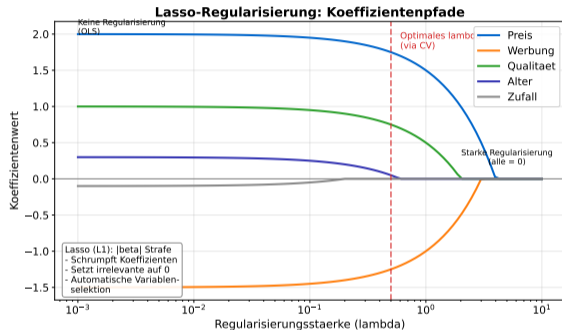
January 25, 2026

Problem:

- Viele Prädiktoren → Overfitting
- Kleine Stichprobe → instabile Schätzungen
- Multikollinearität → hohe Varianzen

Loesung:

- Strafterm λ penalisiert grosse Koeffizienten
- Bias-Variance-Tradeoff: mehr Bias, weniger Varianz
- Lambda via Kreuzvalidierung optimieren



Regularisierung reduziert Overfitting durch Bestrafung grosser Koeffizienten

Zielfunktion:

$$\min_{\beta} \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^p \beta_j^2$$

Eigenschaften:

- **L2-Norm:** Summe der quadrierten Koeffizienten
- **Schrumpfung:** Koeffizienten werden kleiner, aber nicht 0
- **Alle Variablen bleiben im Modell**
- Gut bei Multikollinearität: verteilt Gewicht auf korrelierte Variablen
- $\lambda = 0$: OLS-Regression, $\lambda \rightarrow \infty$: alle $\beta_j \rightarrow 0$

Nachteil: Keine Variablenselektion \rightarrow Modell bleibt komplex

Ridge schrumpft Koeffizienten, setzt sie aber nie auf exakt 0

Zielfunktion:

$$\min_{\beta} \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^p |\beta_j|$$

Eigenschaften:

- **L1-Norm:** Summe der Absolutwerte der Koeffizienten
- **Schrumpfung UND Selektion:** Koeffizienten werden 0
- **Sparse Models:** Nur wichtigste Variablen bleiben
- Automatische Feature Selection
- Bei korrelierten Variablen: wählt eine, setzt andere auf 0

Vorteil: Interpretierbare, einfache Modelle mit wenigen Variablen

Lasso führt Variablenselektion durch – Koeffizienten können exakt 0 werden

Wann Ridge, wann Lasso, wann Elastic Net?

Ridge (L2)

- Viele kleine bis mittlere Effekte
- Multikollinearität: alle Variablen sollen bleiben
- Prognose wichtiger als Interpretation

Lasso (L1)

- Sparsity gewünscht: nur wenige wichtige Variablen
- Interpretation wichtig: einfaches Modell
- Variablenselektion als Nebeneffekt

Elastic Net (Kombination)

- $\alpha \in [0, 1]$: $\alpha = 0$ ist Ridge, $\alpha = 1$ ist Lasso
- Kompromiss: etwas Selektion, aber robuster als Lasso
- Gut bei $p > n$ und Gruppen korrelierter Variablen

Ridge fuer Stabilität, Lasso fuer Sparsity, Elastic Net fuer beides

```
library(glmnet)

# Datenvorbereitung: Matrix X, Vektor y
X <- model.matrix(y ~ . - 1, data = train)
y <- train$y

# Kreuzvalidierung fuer optimales Lambda
cv_lasso <- cv.glmnet(X, y, alpha = 1) # alpha=1: Lasso
cv_ridge <- cv.glmnet(X, y, alpha = 0) # alpha=0: Ridge
cv_enet <- cv.glmnet(X, y, alpha = 0.5) # Elastic Net

# Beste Lambda-Werte
lambda_min <- cv_lasso$lambda.min # minimiert MSE
lambda_1se <- cv_lasso$lambda.1se # sparsamer

# Finales Modell
model <- glmnet(X, y, alpha = 1, lambda = lambda_min)
coef(model) # Koeffizienten (viele = 0 bei Lasso)
```

cv.glmnet findet optimales Lambda via Kreuzvalidierung – alpha steuert L1/L2-Mix