

Modellauswahl & Regularisierung

Wir wollen die wahre Vorhersagegüte realistisch einschätzen, Über/Unteranpassungen vermeiden und die Modellkomplexität bewusst steuern.

Wir nutzen Lern- & Validierungskurven, k-fache Kreuzvalidierung und verschaltete Kreuzvalidierung.

Intuition: Nicht das Modell, das Training am besten auswendig kann, gewinnt – sondern das, das neue Daten gut trifft.

A1. Problemrahmen

- Datenmodell: $\bar{y} = f^*(\bar{x}) + \varepsilon \quad E(\bar{x}) = 0 \quad \text{VAR}(\bar{x}) = \sigma^2$
- Prädiktor: Aus Trainingsdaten D lernen wir $\hat{f}_D(x)$

$$\text{MSE}_{\text{test}} = E[(y - \hat{f}_D(\bar{x}))^2]$$

Intuition: Wie sehr verfehlt sich unser Modell im Durchschnitt bei neuen Fällen?

A2. Bias-Varianz-Zerlegung

Zerlegung: $E[(y - \hat{f}_D(\bar{x}))^2 | \bar{x}] = \sigma^2 + (E(\hat{f}_D(\bar{x})) - f^*(\bar{x}))^2 + \text{VAR}(\hat{f}_D(\bar{x}))$

Zerlegung

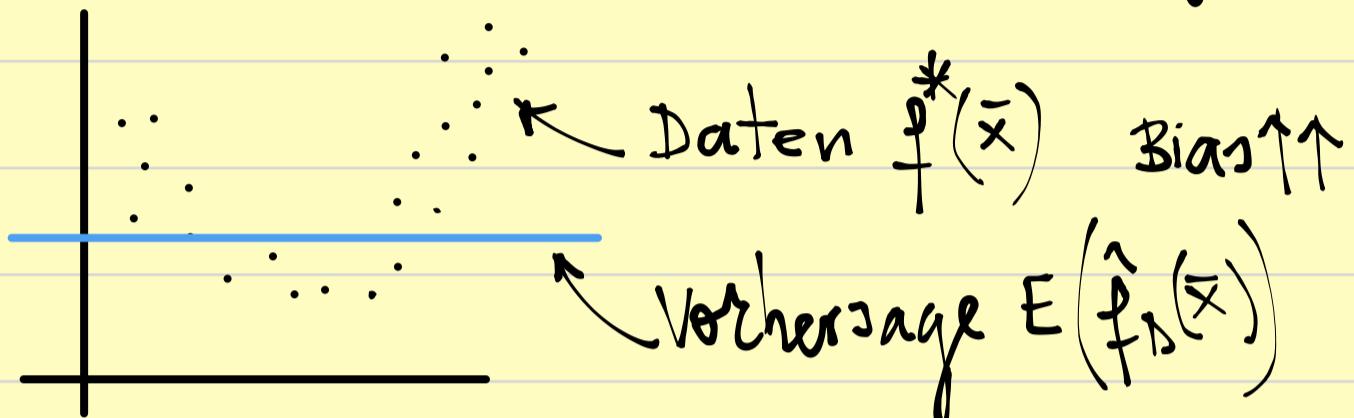
Rauschen

Bias²

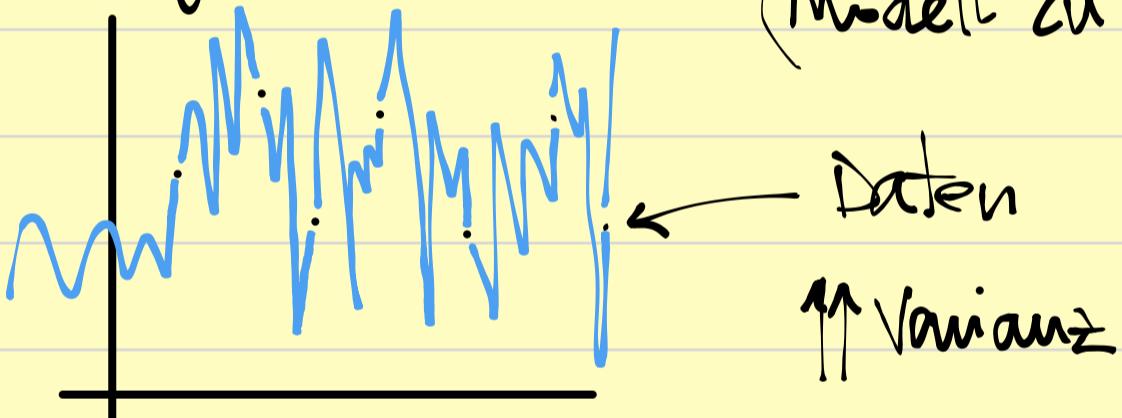
Varianz

Intuition:

BIAS. systematischer Fehler (Modell zu einfach)



VARIANZ. Empfindlichkeit ggü. der Stichprobe
(Modell zu zappelig)



Extrem zackiger Fit, der jeden Punkt trifft.

Hebel: KOMPLEXITÄTSERHÖHUNG : Bias↓ Varianz↑

KOMPLEXITÄTSREDUKTION : Bias↑ Varianz↓

„Der „Sweet-Spot“ liegt dort, wo die Summe aus Bias² + Varianz + Rauschen minimal ist.“

A3. Regularisierung: kontrollierte Einfachheit. Ridge / Lasso.

Ausgangspunkt: $\min_{\mathbf{w}} \frac{1}{n} \|\bar{\mathbf{y}} - \bar{\mathbf{X}}\bar{\mathbf{w}}\|_2^2$

- Ridge-Regression: (L2-Strafe)

Alle Gewichte werden gleichmäßig in Richtung Null gezogen („Summiband“). Gut bei Multikolinearität und wenn alle Merkmale etwas beitragen.

$$\min_{\mathbf{w}} \frac{1}{n} \|\bar{\mathbf{y}} - \bar{\mathbf{X}}\bar{\mathbf{w}}\|_2^2, \quad \hat{\mathbf{w}}_\lambda = (\bar{\mathbf{X}}^T \bar{\mathbf{X}} + n\lambda I)^{-1} \bar{\mathbf{X}}^T \bar{\mathbf{y}}$$

- Lasso: (L1-Strafe)

Viele Gewichte werden genau auf Null gefahren (Sparsamkeit). Lasso macht automatische Merkmalauswahl.

$$\min_{\mathbf{w}} \frac{1}{n} \|\bar{\mathbf{y}} - \bar{\mathbf{X}}\bar{\mathbf{w}}\|_2^2 + \lambda \|\bar{\mathbf{w}}\|_1$$

Wahlhilfe: Ridge(L2) bei vielen schwach-nützlichen Merkmalen.

Lasso(L1) wenn nur wenige wirklich relevant sind.

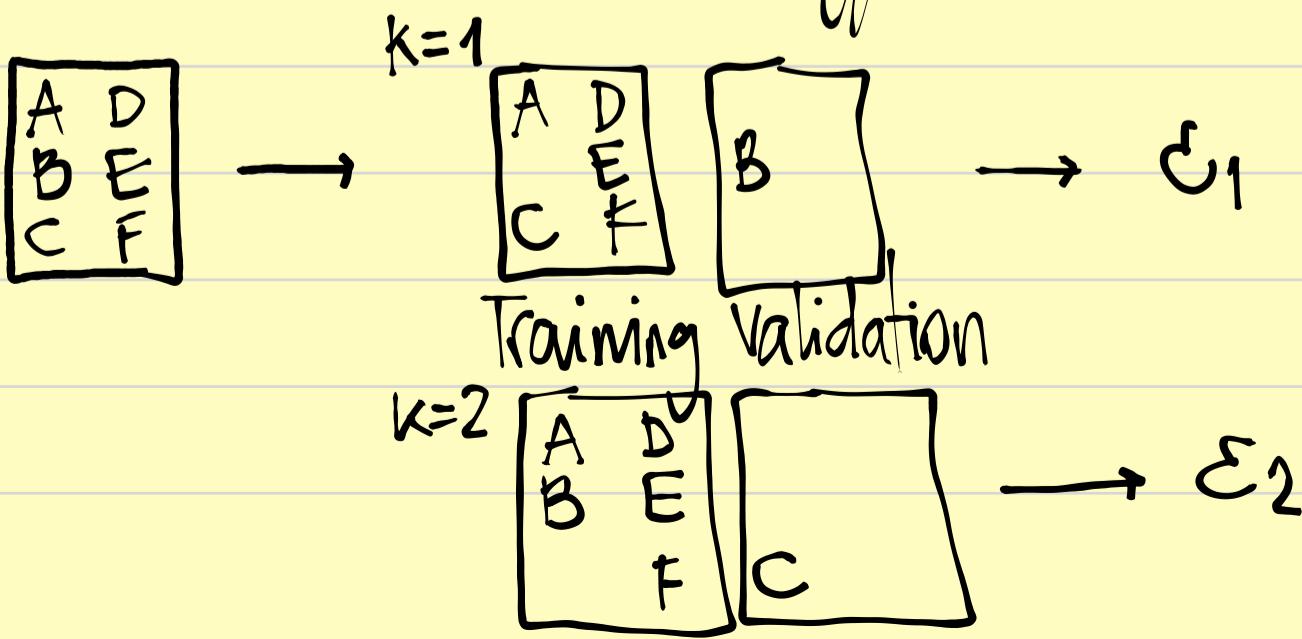
Erweiterung: „Elastic Net“ kombiniert L1+L2 robust bei stark kontrollierten Merkmalen.

λ : ist der .. KOMPLEXITÄTSKNOPF ..

- klein : flexibel (Varianz \uparrow)
- groß : glatt (Bias \uparrow)

2.5. Train / Validation / Test bzw K-fache Kreuzvalidierung

- Ablauf: Train (lernen)
Validation (Hyperparameter Anpassung)
Test (einzahlige Endmessung)
- K-fache Kreuzvalidierung: die Daten werden in K Gruppen / Falten getrennt; wiederholt auf $K-1$ Falten trainiert, auf der verbleibenden validiert, und alle Ergebnisse gemittelt.
- Praxis: CV (Crossvalidation) verringert Zufallschwankung. liefert stabilere Schätzungen und nutzt Daten effizienter.



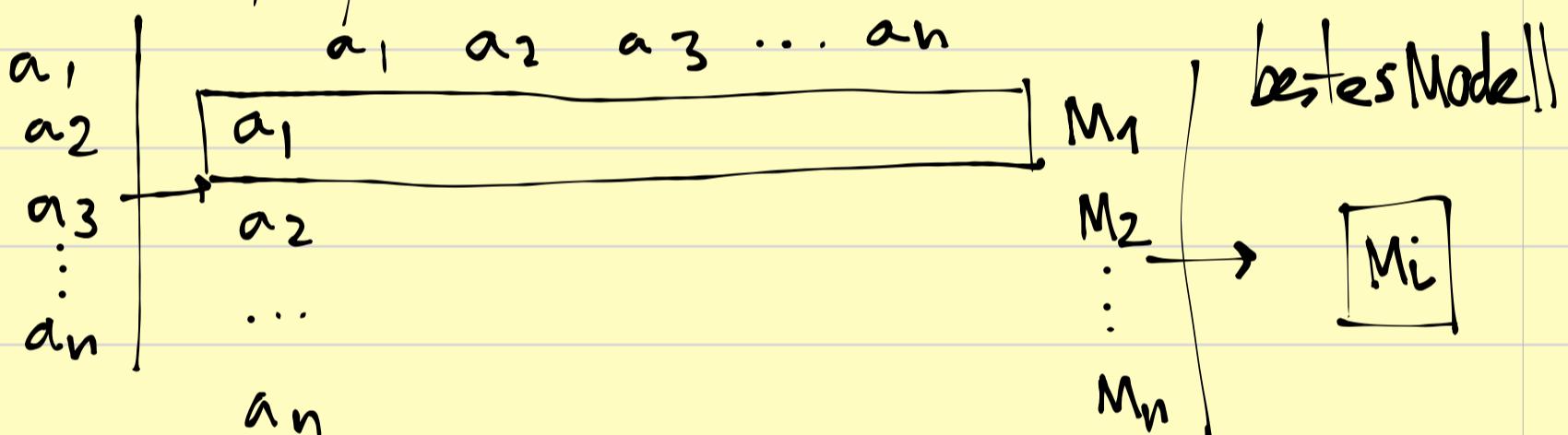
Training Validation

...

$\bar{\epsilon}$

16. Hypersparameter Suche (Grid/Random) & Nested CV.

- Grid-Suche: Systematisch über ein Raster werden Hyperparameter untersucht.



- Zufallsuche: zufällige Kombinationen. wird in höher Dimensionen effizienter.
- Nested. CV: Tuning (Grid/Random) der HP per CV.

Checkliste:

- 1) Daten in Train/Validation/Test teilen
- 2) Pipeline bauen
- 3) Lernkurve prüfen.

- 4) Validierungskurve per CV fold.
 λ /Tiefe wählen (Grid/Random)
- 5) Endmodell auf Train fitten und einmal Test auswerten.
- 6) Bericht mit Zahlen & Plots erstellen
ggf. Entscheidung über HP begründen
(Bias, Varianz).

