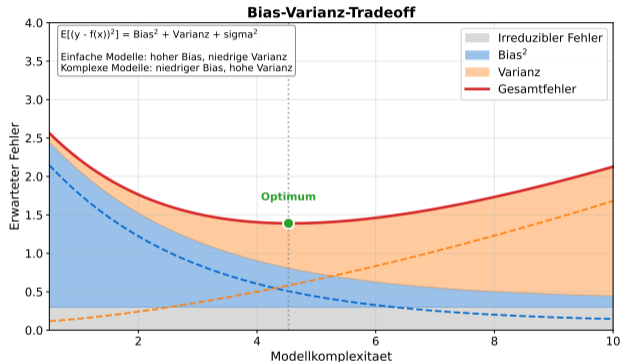


## Topic 3: Bias-Varianz-Tradeoff

Block 3: Komplexitaet, Kausalitaet & Generalisierung

January 25, 2026

- Jedes ML-Modell hat zwei Fehlerquellen:
  - **Bias**: Systematische Abweichung
  - **Varianz**: Zufällige Schwankung
- Beide haengen von Modellkomplexitaet ab
- Zielkonflikt: Reduktion des einen erhoht oft das andere
- Optimale Komplexitaet minimiert Gesamtfehler



Der Bias-Varianz-Tradeoff bestimmt die optimale Modellkomplexitaet

## Definition

Bias misst die **systematische Abweichung** zwischen durchschnittlicher Vorhersage und wahrem Wert

### Ursachen fuer hohen Bias:

- Modell ist **zu einfach** (z.B. lineare Regression fuer nichtlineare Daten)
- Wichtige Zusammenhaenge werden **nicht erfasst**
- Zu wenige Features oder zu starke Einschränkungen
- Fuehrt zu **Underfitting**

### Konsequenz:

- Modell ist sowohl auf Trainings- als auch Testdaten schlecht
- Fehler ist konsistent hoch, auch bei mehr Daten

---

Hoher Bias bedeutet: Das Modell ist zu simpel fuer die Daten

## Definition

Varianz misst die **Streuung der Vorhersagen** bei unterschiedlichen Trainingsdaten

### Ursachen fuer hohe Varianz:

- Modell ist **zu komplex** (z.B. Polynom hohen Grades)
- Lernt **Rauschen** statt wahre Muster
- Zu viele Parameter im Verhaeltnis zu Datenpunkten
- Fuehrt zu **Overfitting**

### Konsequenz:

- Modell passt perfekt auf Trainingsdaten
- Schlechte Performance auf neuen, ungesehenen Daten
- Vorhersagen variieren stark bei leicht veraenderten Trainingsdaten

---

Hohe Varianz bedeutet: Das Modell ist zu flexibel und lernt Rauschen

## Mathematische Zerlegung

$$\mathbb{E} [(y - \hat{y})^2] = \text{Bias}(\hat{y})^2 + \text{Var}(\hat{y}) + \sigma^2$$

### Komponenten:

- $\text{Bias}^2$ : Quadrierte systematische Abweichung
- $\text{Var}$ : Varianz der Vorhersagen
- $\sigma^2$ : Irreduzibler Fehler (Rauschen in Daten)

### Der Tradeoff:

- **Einfaches Modell**: Hoher Bias, niedrige Varianz
- **Komplexes Modell**: Niedriger Bias, hohe Varianz
- **Optimales Modell**: Balanciert beide Fehlerquellen

---

Optimale Modellkomplexität minimiert  $\text{Bias}^2 + \text{Varianz}$

## Symptome von Overfitting (hohe Varianz):

- **Trainingsfehler** sehr klein
- **Testfehler** deutlich groesser
- Grosse Luecke zwischen Training- und Testperformance

## Gegenmassnahmen:

1. **Mehr Trainingsdaten** sammeln
2. **Feature Selection**: Irrelevante Features entfernen
3. **Regularisierung**: L1 (Lasso) oder L2 (Ridge) Penalties
4. **Cross-Validation**: Robuste Modellwahl
5. **Ensemble-Methoden**: Bagging, Random Forests
6. **Early Stopping**: Training vorzeitig beenden

---

Cross-Validation und Regularisierung sind die wichtigsten Werkzeuge gegen Overfitting