

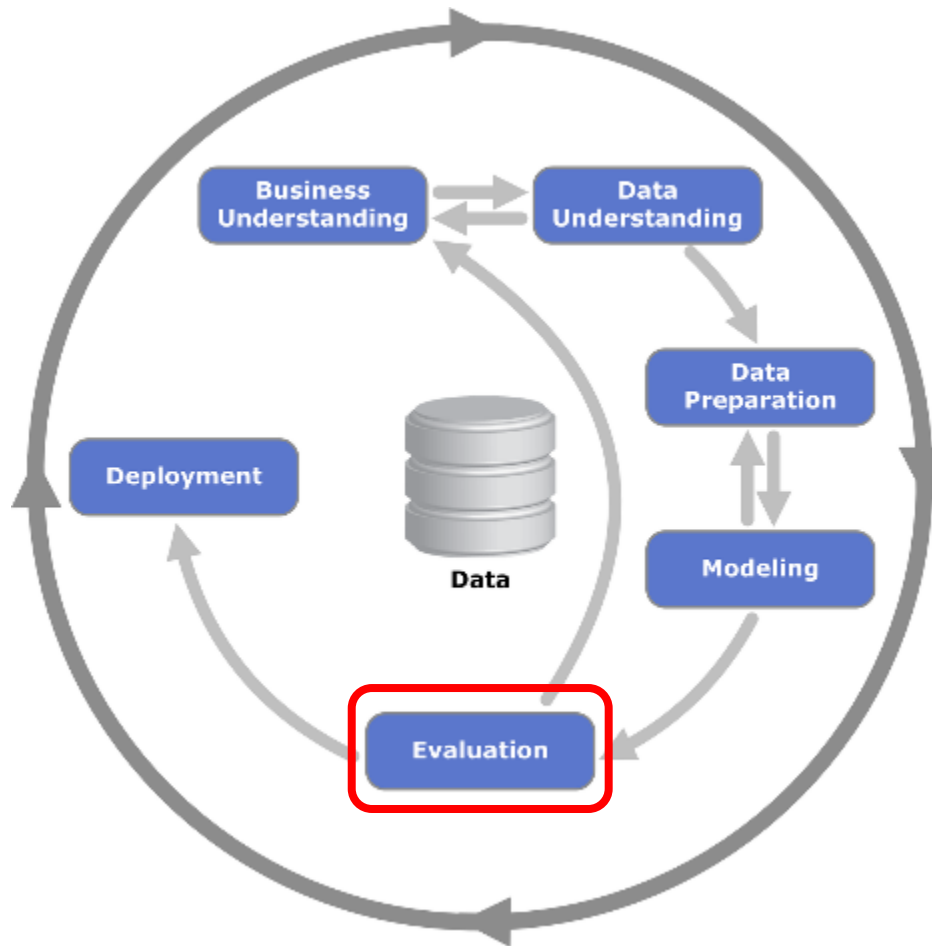


FAKULTÄT FÜR  
INFORMATIK

# Model Validation

Alexander Dockhorn

# Was machen wir hier eigentlich?



# Bewertung von Klassifikatoren

Klassifikationsmodelle können anhand ihrer „Genauigkeit“ bewertet werden

- wie ist Genauigkeit definiert?
  - Bewertungsmaße
- wie können wir unterschiedliche Modelle fair vergleichen?
  - Validierungsverfahren
- Was beeinflusst die Genauigkeit unseres Modells?
  - Bias–Varianzproblem

# Inhalt

## 1. Bewertungsmaße

1. Confusion Matrix
2. Accuracy
3. Precision
4. Recall
5. F1-Measure

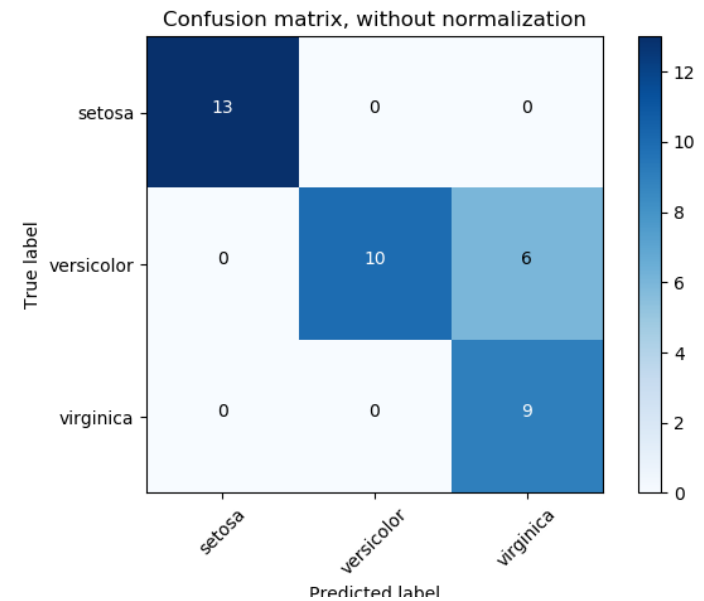
## 2. Validierungsverfahren

1. Simple Validation
2. Crossvalidation
3. Train-Val-Test Split

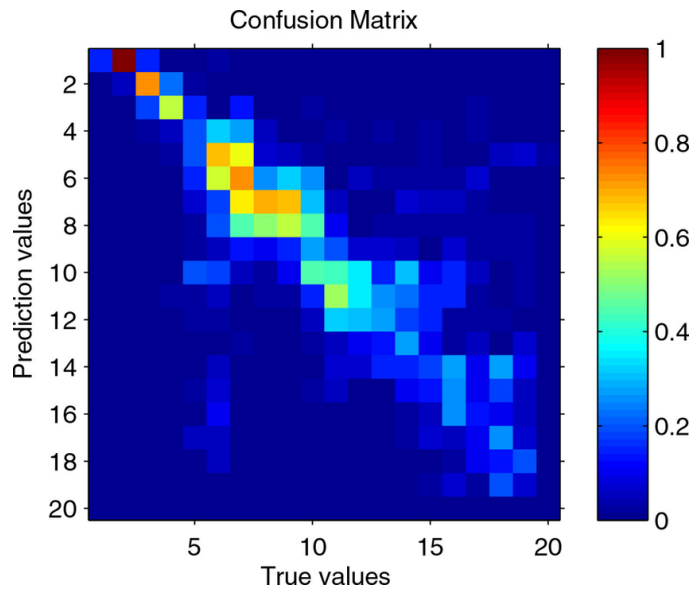
## 3. Bias-vs.-Varianzproblem

# Confusion Matrix

	$p'$ (Predicted)	$n'$ (Predicted)
$P$ (Actual)	True Positive	False Negative
$n$ (Actual)	False Positive	True Negative



# Confusion Matrix



215	0	2	0	5	2	224
0	135	34	0	2	40	211
0	16	368	1	0	12	397
1	0	2	458	0	0	461
3	0	1	20	183	30	237
0	36	12	0	8	414	470
219	187	419	479	198	498	
						predicted class

# Bewertungsmaße

- Accuracy

$$\frac{TP + TN}{TP + TN + FP + FN}$$

- Precision

$$\frac{TP}{TP + FP}$$

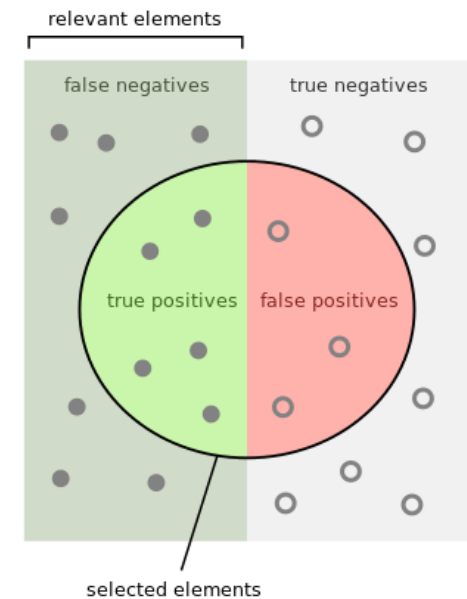
- Recall

$$\frac{TP}{TP + FN}$$

- F1-Measure

$$\frac{2 \cdot \textit{precision} \cdot \textit{recall}}{\textit{precision} + \textit{recall}}$$

	p' (Predicted)	n' (Predicted)
P (Actual)	True Positive	False Negative
n (Actual)	False Positive	True Negative



How many selected items are relevant?

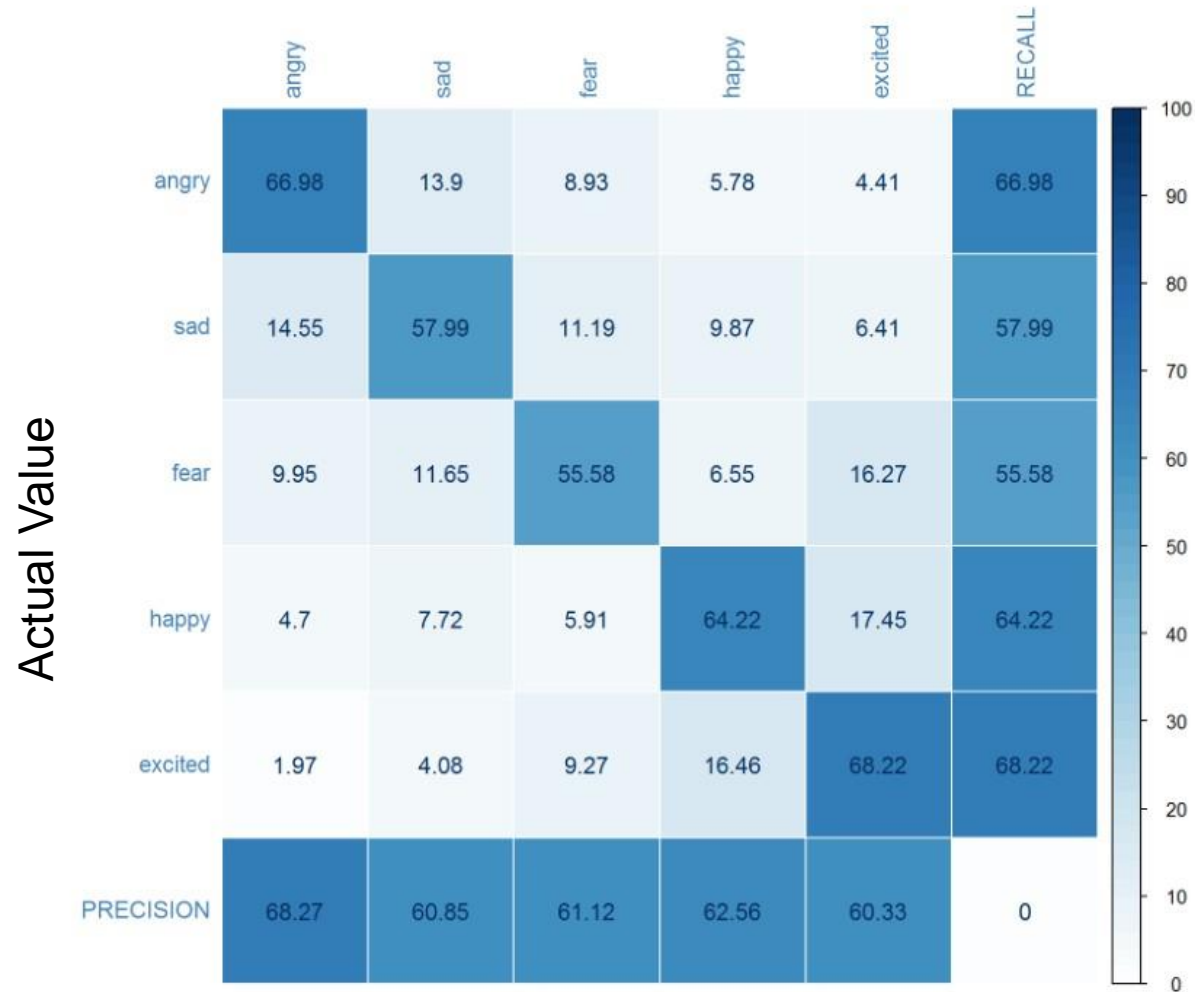
$$\textit{Precision} = \frac{\text{green}}{\text{green} + \text{red}}$$

How many relevant items are selected?

$$\textit{Recall} = \frac{\text{green}}{\text{green}}$$

# Confusion Matrix

Predicted Value





# Validierungsverfahren

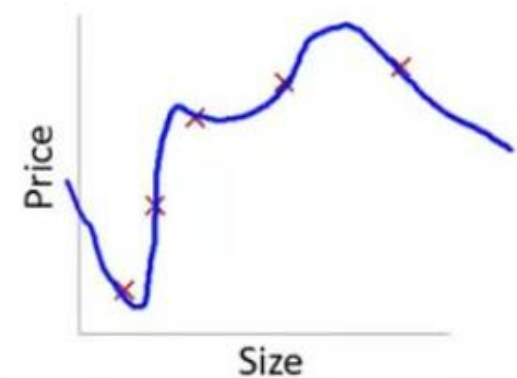
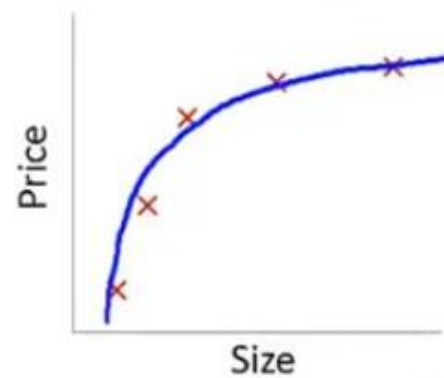
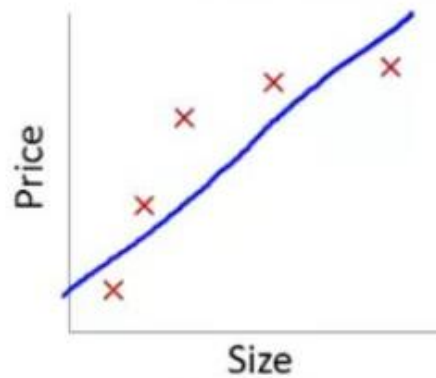
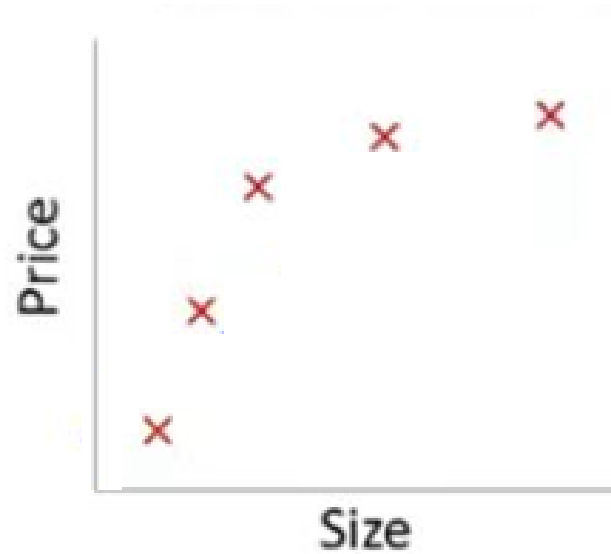
## Wie können wir einen Klassifikator bewerten?

- Aufteilen des Datensatzes in Trainings- und Testdaten
- Klassifikator auf Trainingsdaten trainieren
- Und Accuracy oder F1-Wert auf Testdaten bestimmen

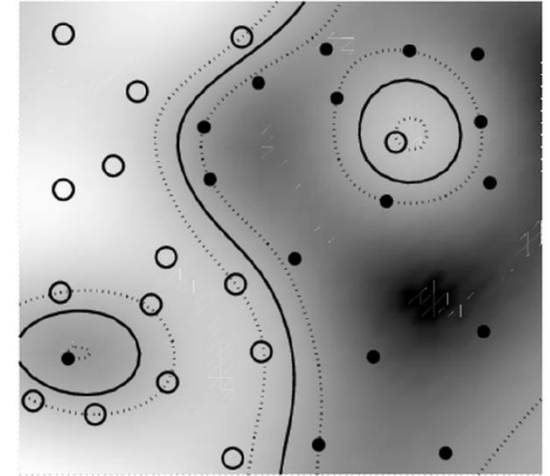
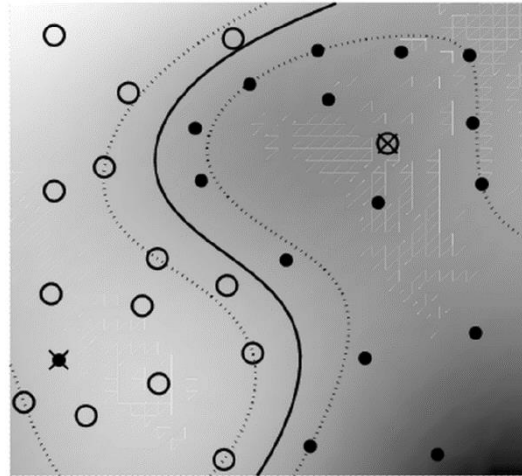
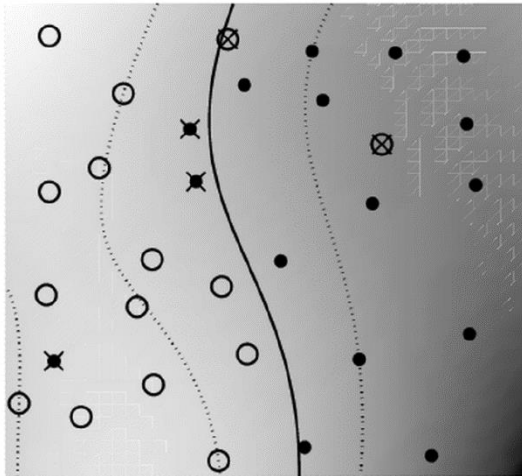
## Was können wir machen, wenn wir nur wenige Daten haben?

- Aufteilen des Datensatzes in mehrere Gruppen
- Nimm jeweils einige der Gruppen als Trainingsdatensatz
- ...und die übrigen Gruppen als Testdatensatz
  
- Wiederhole das Vorgehen über unterschiedliche Train/Test-Aufteilungen und middle das Ergebnis

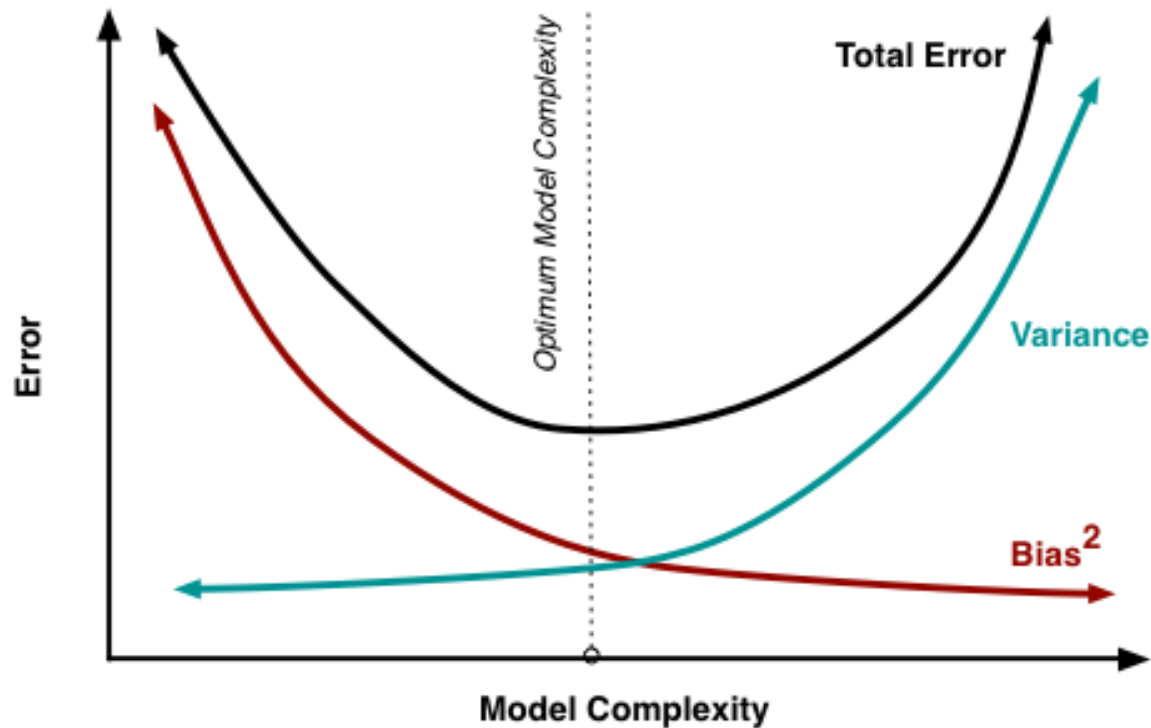
# Overfitting vs. Underfitting in Regression



# Overfitting vs. Underfitting in Classification



# Overfitting vs. Underfitting oder auch Bias and Variance



# Bekämpfen des Bias und Variance Problems

## Was können wir dagegen machen?

- Problemklasse feststellen
  - Train- und Testfehler beobachten
- Bei einem Bias Problem:
  - Weitere Daten hinzufügen
  - Komplexität durch Parametereinstellungen erhöhen
- Bei einem Variance Problem:
  - Attribute und Ausreißer entfernen
  - Komplexität durch Parametereinstellung reduzieren