Maschinelles Lernen zur Modellbildung in den Naturwissenschaften

Niels Landwehr

Institut für Informatik, Universität Potsdam

Vorstellung Forschungsgruppe

- Nachwuchsgruppe „Machine Learning and Scientific Data Analysis“
  - Emmy Noether-Programm der DFG, seit Februar 2013
  - Forschungsprogramm: Fragestellungen des maschinellen Lernens im Zusammenhang mit naturwissenschaftlicher Modellbildung
  - Kooperation mit Wissenschaftlern aus der Geophysik (AG Frank Scherbaum) und der kognitiven Psychologie (AG Reinhold Kliegl)
  - Angegliedert an die Arbeitsgruppe „Maschinelles Lernen“ (Tobias Scheffer), Institut für Informatik, Universität Potsdam

- Heute: Überblick über Forschungsprogramm und Vorarbeiten

Modellbildung aus Beobachtungsdaten

Natürliches Phänomen: Erdbeben

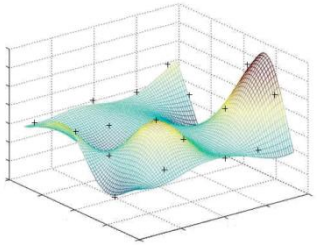
Beobachtung

Daten

ID	D	F	M	R	I	H	PGA
1	100	305	4,50	2,12	0	3,15	2,12
2	122	400	3,10	1,75	0	2,05	1,75
3	231	265	5,40	2,45	1	3,54	2,45
4	146	300	3,80	1,93	1	1,32	1,93
5	133	195	4,90	2,44	1	2,44	2,44
6	215	245	3,20	1,63	0	1,33	1,63
7	100	305	4,50	2,12	1	1,33	2,12
8	122	400	3,10	1,75	0	1,98	1,75
9	231	265	5,40	2,45	1	2,05	2,45
10	146	300	3,80	1,93	1	3,54	1,93
11	133	195	4,90	2,44	1	1,32	2,44
12	215	245	3,20	1,63	0	2,44	1,63
13	251	310	6,50	2,83	1	1,33	2,83
14	133	195	4,90	2,44	1	1,33	2,44
15	215	245	3,20	1,63	1	3,21	1,63

Modellbildung

Modell



Beschreibung, Vorhersage

Modellbildung aus Beobachtungsdaten

Natürliches Phänomen: Erdbeben

Installation von Beschleunigungssensoren.

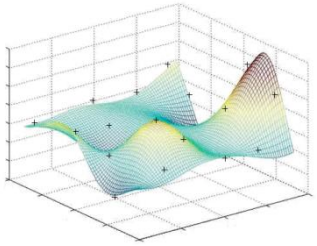
Beobachtung

Daten

ID	D	F	M	R	I	H	PGA
1	100	305	4,50	2,12	0	3,15	2,12
2	122	400	3,10	1,75	0	2,05	1,75
3	231	265	5,40	2,45	1	3,54	2,45
4	146	300	3,80	1,93	1	1,32	1,93
5	133	195	4,90	2,44	1	2,44	2,44
6	215	245	3,20	1,63	0	1,33	1,63
7	100	305	4,50	2,12	1	1,33	2,12
8	122	400	3,10	1,75	0	1,98	1,75
9	231	265	5,40	2,45	1	2,05	2,45
10	146	300	3,80	1,93	1	3,54	1,93
11	133	195	4,90	2,44	1	1,32	2,44
12	215	245	3,20	1,63	0	2,44	1,63
13	251	310	6,50	2,83	1	1,33	2,83
14	133	195	4,90	2,44	1	1,33	2,44
15	215	245	3,20	1,63	1	3,21	1,63

Modellbildung

Modell



Beschreibung, Vorhersage

Modellbildung aus Beobachtungsdaten

Lernen eines Vorhersagemodells:

$$y^* = \arg\max_y p(y|\mathbf{x};\theta).$$

Bodenbeschleunigung

Magnitude, Entfernung zum Epizentrum, Herdmechanismus, …

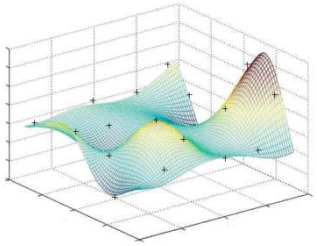
Daten

ID	D	F	M	R	I	H	PGA
1	100	305	4,50	2,12	0	3,15	2,12
2	122	400	3,10	1,75	0	2,05	1,75
3	231	265	5,40	2,45	1	3,54	2,45
4	146	300	3,80	1,93	1	1,32	1,93
5	133	195	4,90	2,44	1	2,44	2,44
6	215	245	3,20	1,63	0	1,33	1,63
7	100	305	4,50	2,12	1	1,33	2,12
8	122	400	3,10	1,75	0	1,98	1,75
9	231	265	5,40	2,45	1	2,05	2,45
10	146	300	3,80	1,93	1	3,54	1,93
11	133	195	4,90	2,44	1	1,32	2,44
12	215	245	3,20	1,63	0	2,44	1,63
13	251	310	6,50	2,83	1	1,33	2,83
14	133	195	4,90	2,44	1	1,33	2,44
15	215	245	3,20	1,63	1	3,21	1,63

Beschreibung, Vorhersage

Modellbildung

Modell



Modellbildung aus Beobachtungsdaten

- Einsicht in natürliche Prozesse.
- Vorhersagen lokaler Bodenbeschleunigung: seismische Risikoanalyse, Auslegung von Gebäuden.

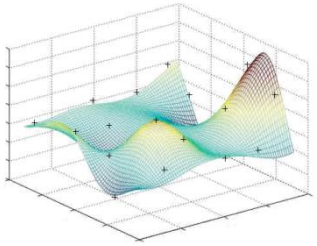
Natürliches Phänomen: Erdbeb

ntung

ID	D	F	M	R	I	H	PGA
1	100	305	4,50	2,12	0	3,15	2,12
2	122	400	3,10	1,75	0	2,05	1,75
3	231	265	5,40	2,45	1	3,54	2,45
4	146	300	3,80	1,93	1	1,32	1,93
5	133	195	4,90	2,44	1	2,44	2,44
6	215	245	3,20	1,63	0	1,33	1,63
7	100	305	4,50	2,12	1	1,33	2,12
8	122	400	3,10	1,75	0	1,98	1,75
9	231	265	5,40	2,45	1	2,05	2,45
10	146	300	3,80	1,93	1	3,54	1,93
11	133	195	4,90	2,44	1	1,32	2,44
12	215	245	3,20	1,63	0	2,44	1,63
13	251	310	6,50	2,83	1	1,33	2,83
14	133	195	4,90	2,44	1	1,33	2,44
15	215	245	3,20	1,63	1	3,21	1,63

Beschreibung, Vorhersage

Modell



Modellbildung

Modellbildung aus Experimentaldaten

Natürliches Phänomen: Blickbewegungen

Aufnahme der Blickbewegung im Labor.

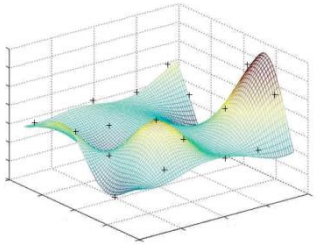
Beobachtung

Daten

ID	D	F	M	R	I	H	PGA
1	100	305	4,50	2,12	0	3,15	2,12
2	122	400	3,10	1,75	0	2,05	1,75
3	231	265	5,40	2,45	1	3,54	2,45
4	146	300	3,80	1,93	1	1,32	1,93
5	133	195	4,90	2,44	1	2,44	2,44
6	215	245	3,20	1,63	0	1,33	1,63
7	100	305	4,50	2,12	1	1,33	2,12
8	122	400	3,10	1,75	0	1,98	1,75
9	231	265	5,40	2,45	1	2,05	2,45
10	146	300	3,80	1,93	1	3,54	1,93
11	133	195	4,90	2,44	1	1,32	2,44
12	215	245	3,20	1,63	0	2,44	1,63
13	251	310	6,50	2,83	1	1,33	2,83
14	133	195	4,90	2,44	1	1,33	2,44
15	215	245	3,20	1,63	1	3,21	1,63

Modellbildung

Modell



Beschreibung, Vorhersage

Modellbildung aus Experimentaldaten

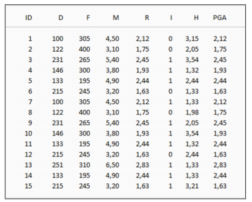
Lernen eines Vorhersagemodells:

$$y^* = \arg\max_y p(y|\mathbf{x};\theta).$$

Blickbewegung

Syntax und Semantik des Textes

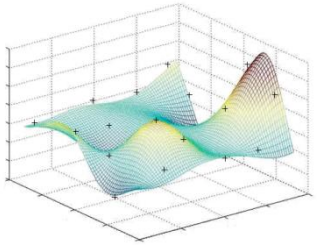
Daten



Modellbildung

Beschreibung, Vorhersage

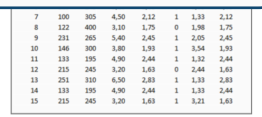
Modell



Modellbildung aus Experimentaldaten

- Blickbewegungen zentrales Forschungsobjekt zum Verständnis des Zusammenspiels von Kognition und Motorsteuerung.
- Biometrische Anwendungen.

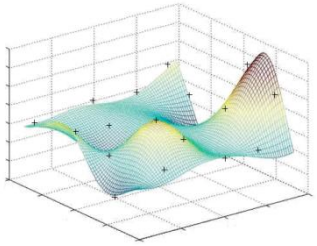
Natürliches Phänomen: Blickbe



Beschreibung, Vorhersage

Modellbildung

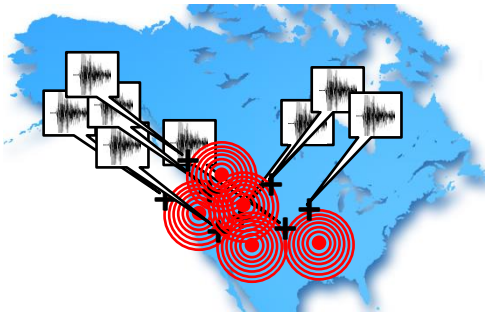
Modell



Maschinelles Lernen

- Maschinelles Lernen: Schätzen von Vorhersagemodellen aus Beobachtungen

Beobachtungen

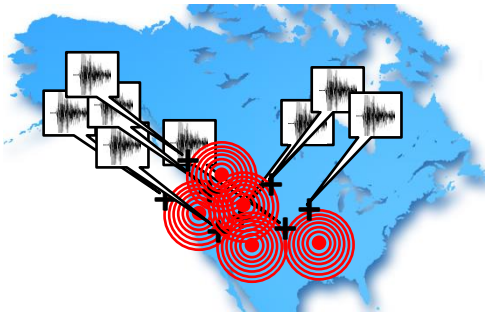


Vorhersage

Maschinelles Lernen

- Maschinelles Lernen: Schätzen von Vorhersagemodellen aus Beobachtungen

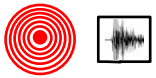
Beobachtungen



Vorhersage

Unabhängige Variablen: beschreiben Erdbeben

Abhängige Variable: Bodenbeschleunigung



$(\mathbf{x}_1, y_1)$

$(\mathbf{x}_2, y_2)$

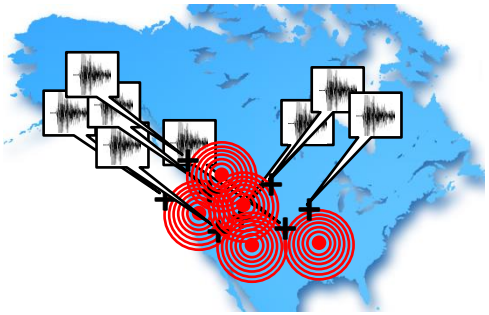
...

$(\mathbf{x}_n, y_n)$

Maschinelles Lernen

- Maschinelles Lernen: Schätzen von Vorhersagemodellen aus Beobachtungen

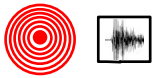
Beobachtungen



Vorhersage

Unabhängige Variablen: beschreiben Erdbeben

Abhängige Variable: Bodenbeschleunigung



$(\mathbf{x}_1, y_1)$

$(\mathbf{x}_2, y_2)$

...

$(\mathbf{x}_n, y_n)$

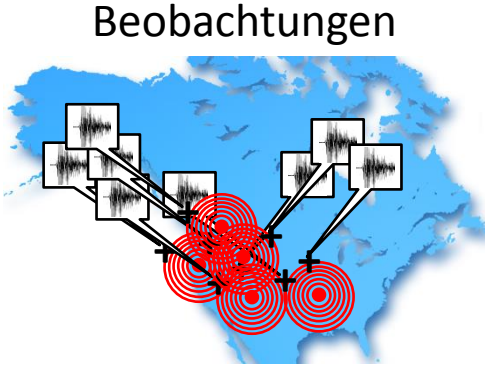
Vorhersagemodell

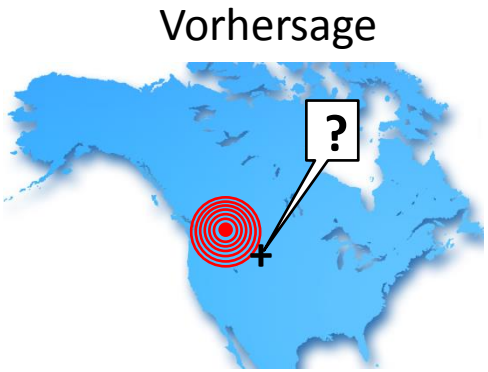
$$f_\theta : \mathcal{X} \rightarrow \mathcal{Y}$$

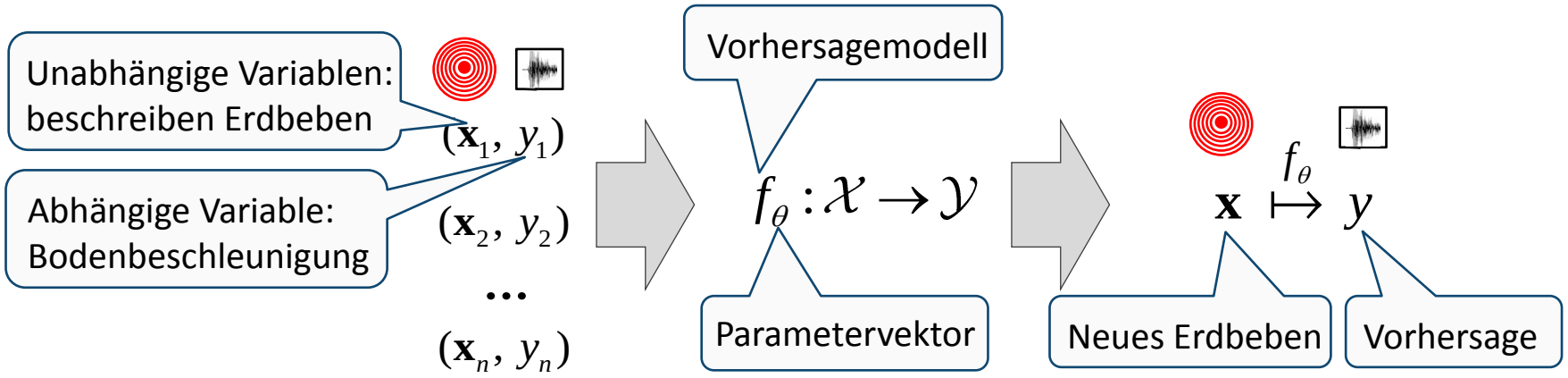
Parametervektor

Maschinelles Lernen

- Maschinelles Lernen: Schätzen von Vorhersagemodellen aus Beobachtungen







Maschinelles Lernen: Modellinferenz

- Zentraler Schritt: Inferenz eines geeigneten Modells gegeben Daten

- Viele Modellklassen und Verfahren
  - Probabilistische Modelle mit max likelihood /max aposteriori / Bayes‘scher Inferenz
  - parametrische / nichtparametrische (kernelisierte) Modelle

- Lösen eines Optimierungsproblems
  - Suchproblem über komplexen Raum möglicher Modelle
  - Algorithmisch aufwendig
  - Approximative numerische Verfahren

Modellbildung: statistische Annahmen

- **I.I.D. Annahme**:
  - Alle Datenpunkte folgen einer einheitlichen Verteilung $p(\mathbf{x}, y)$.
  - Datenpunkte sind unabhängige Stichproben aus $p(\mathbf{x}, y)$.

Natürliches Phänomen: Erdbeben

Beobachtung

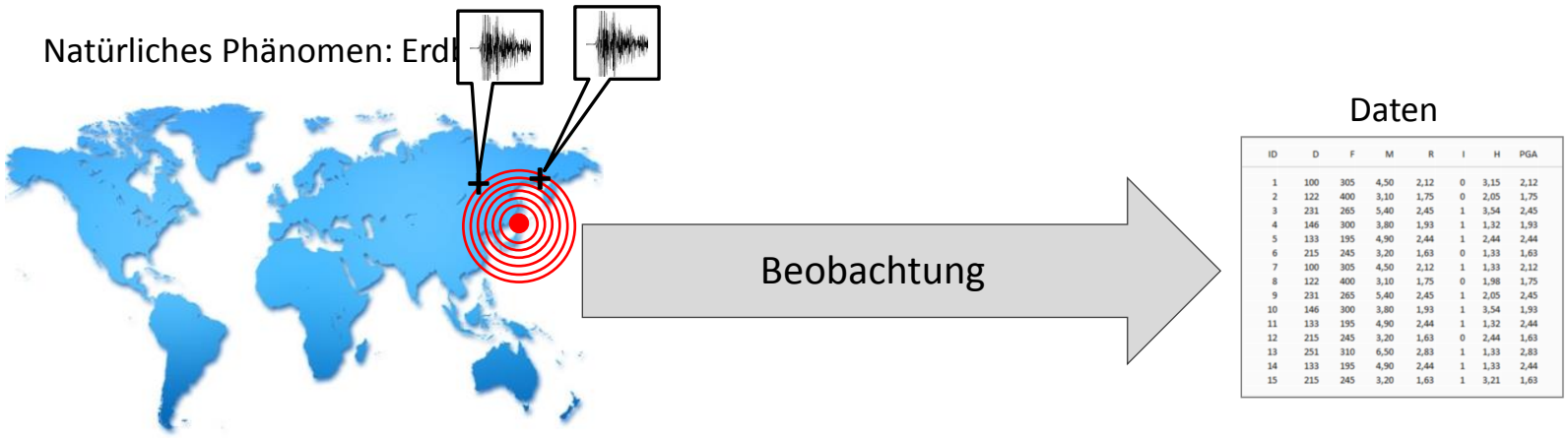
Daten

ID	D	F	M	R	I	H	PGA
1	100	305	4,50	2,12	0	3,15	2,12
2	122	400	3,10	1,75	0	2,05	1,75
3	231	265	5,40	2,45	1	3,54	2,45
4	146	300	3,80	1,93	1	1,32	1,93
5	133	195	4,90	2,44	1	2,44	2,44
6	215	245	3,20	1,63	0	1,33	1,63
7	100	305	4,50	2,12	1	1,33	2,12
8	122	400	3,10	1,75	0	1,98	1,75
9	231	265	5,40	2,45	1	2,05	2,45
10	146	300	3,80	1,93	1	3,54	1,93
11	133	195	4,90	2,44	1	1,32	2,44
12	215	245	3,20	1,63	0	2,44	1,63
13	251	310	6,50	2,83	1	1,33	2,83
14	133	195	4,90	2,44	1	1,33	2,44
15	215	245	3,20	1,63	1	3,21	1,63

Abhängige Datenpunkte

- **Abhängige Datenpunkte**: Verschiedene Sensoren registrieren dasselbe Ereignis (beispielsweise verschiedene Messstationen eine seismische Welle).

Widerspruch zur **Unabhängigkeitsannahme**.

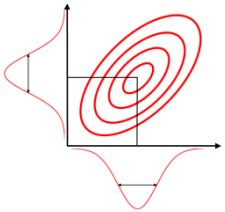


ID	D	F	M	R	I	H	PGA
1	100	305	4,50	2,12	0	3,15	2,12
2	122	400	3,10	1,75	0	2,05	1,75
3	231	265	5,40	2,45	1	3,54	2,45
4	146	300	3,80	1,93	1	1,32	1,93
5	133	195	4,90	2,44	1	2,44	2,44
6	215	245	3,20	1,63	0	1,33	1,63
7	100	305	4,50	2,12	1	1,33	2,12
8	122	400	3,10	1,75	0	1,98	1,75
9	231	265	5,40	2,45	1	2,05	2,45
10	146	300	3,80	1,93	1	3,54	1,93
11	133	195	4,90	2,44	1	1,32	2,44
12	215	245	3,20	1,63	0	2,44	1,63
13	251	310	6,50	2,83	1	1,33	2,83
14	133	195	4,90	2,44	1	1,33	2,44
15	215	245	3,20	1,63	1	3,21	1,63

Abhängige Datenpunkte

- Nicht berücksichtigte Datenabhängigkeiten führen zu einer Unterschätzung der nach der Modellbildung verbleibenden Unsicherheit.

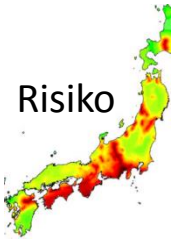
A-posteriori Verteilung über Modellparameter



$$p(y \geq \bar{y}) = \iiint p(y \geq \bar{y}|d, m; \theta) p(\theta|D) p(r, m) d\theta dr dm$$

Seismische Risikoanalyse

Risiko

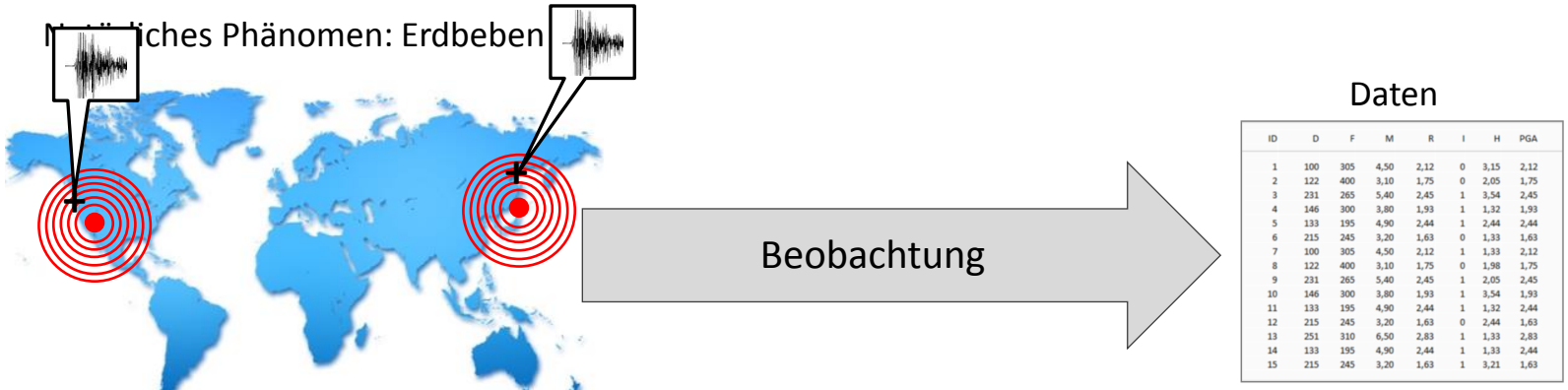


- Falsche Schlussfolgerungen: beispielsweise Unterschätzung der Wahrscheinlichkeit extremer Ereignisse.

Verteilungsverschiebungen

- **Verteilungsverschiebungen**: Verteilung der Daten abhängig vom Ort der Messung (Aufnahme seismischer Daten in unterschiedlichen Regionen).

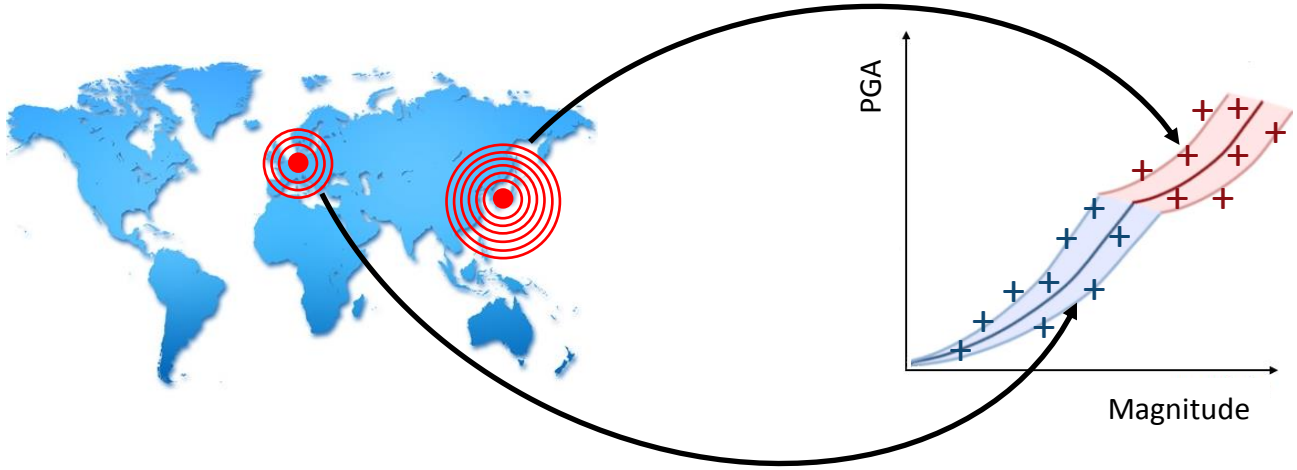
Widerspruch zur **Annahme einer einheitlichen Datenverteilung**.



ID	D	F	M	R	I	H	PGA
1	100	305	4,50	2,12	0	3,15	2,12
2	122	400	3,10	1,75	0	2,05	1,75
3	231	265	5,40	2,45	1	3,54	2,45
4	146	300	3,80	1,93	1	1,32	1,93
5	133	195	4,90	2,44	1	2,44	2,44
6	215	245	3,20	1,63	0	1,33	1,63
7	100	305	4,50	2,12	1	1,33	2,12
8	122	400	3,10	1,75	0	1,98	1,75
9	231	265	5,40	2,45	1	2,05	2,45
10	146	300	3,80	1,93	1	3,54	1,93
11	133	195	4,90	2,44	1	1,32	2,44
12	215	245	3,20	1,63	0	2,44	1,63
13	251	310	6,50	2,83	1	1,33	2,83
14	133	195	4,90	2,44	1	1,33	2,44
15	215	245	3,20	1,63	1	3,21	1,63

Verteilungsverschiebungen

- Bildung von Modellen für seismisches Risiko in Deutschland:
  - In Deutschland stehen nur Beobachtungen niedriger Intensität zur Verfügung.
  - Extrapolation zu Erdbeben höherer Intensität mit Daten aus anderen Regionen.



- Verteilungsunterschiede können zu falschen Vorhersagen führen.

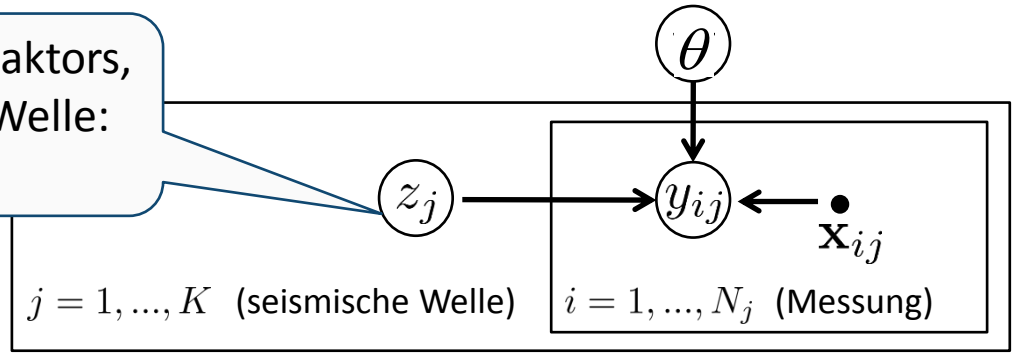
Übersicht

- Einleitung & Motivation
- Lernen unter Datenabhängigkeiten
- Lernen unter Verteilungsverschiebungen
- Modellevaluierung

Abhängigkeiten: Stand der Forschung

- Stand der Forschung zur Modellierung abhängiger Daten: *Mixed Models.*

Einfluss des gemeinsamen Faktors, beispielsweise seismischer Welle: versteckte Variable.



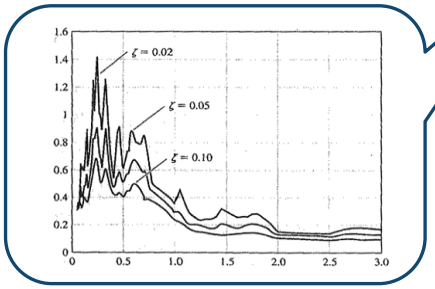
$$\theta^* = \arg\max_{\theta} \prod_{j=1}^{K} \left(\int \left(\prod_{i=1}^{N_j} p(y_{ij}|z_j, \mathbf{x}_{ij}; \theta) \right) p(z_j) dz_j \right)$$

- Bei der Parameterschätzung wird über die versteckten Variablen integriert

Abhängigkeiten: Ziele

- Mixed Models nur bekannt für einfache Ausgabevariablen, in Anwendungen ist Ausgabevariable aber oft komplex strukturiert.

Seismisches Modell: $y^* = \arg\max_y p(y|\mathbf{x};\theta)$



Ausgabe: Antwortspektrum
(Intensität über Frequenz)

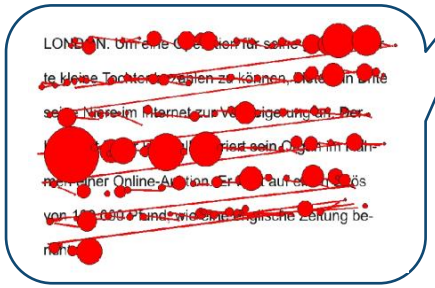
Eingabe: Attribute
des Erdbebens

- **Ziel**: Modelle für komplexe Ausgaben unter Datenabhängigkeiten.
- Aufwendige Optimierungsprobleme bei Modellbildung und Vorhersage.

Abhängigkeiten: Ziele

- Mixed Models nur bekannt für einfache Ausgabevariablen, in Anwendungen ist Ausgabevariable aber oft komplex strukturiert.

Blickbewegungsmodell: $y^* = \arg\max_y p(y|\mathbf{x};\theta)$



Ausgabe: Blickbewegung

LONDON. Um eine Operation für seine erkrankte kleine Tochter zahlen zu können, bot ein Brite seine Niere im Internet zur Versteigerung an. Der verzweifelte Vater offeriert sein Organ im Rahmen einer Online-Auktion. Er hoffte auf einen Erlös von 100.000 Pfund, wie eine englische Zeitung berichtete.

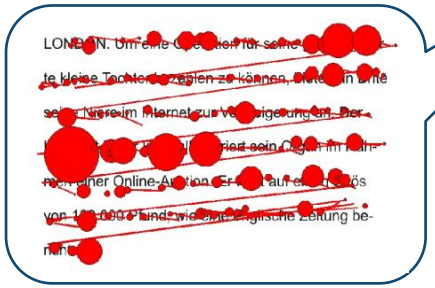
Eingabe: Text

- **Ziel**: Modelle für komplexe Ausgaben unter Datenabhängigkeiten.
- Aufwendige Optimierungsprobleme bei Modellbildung und Vorhersage.

Blickbewegungsmodelle

- **Ziel**: gemeinsame Modelle von Blickbewegungen, Textinhalt und Attributen des Lesers.

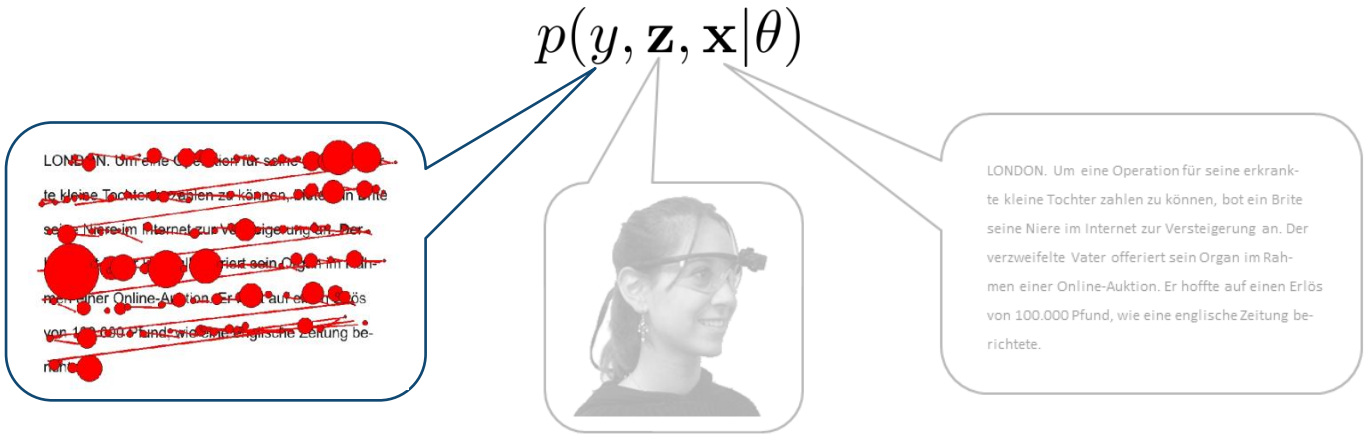
$$p(y, \mathbf{z}, \mathbf{x}|\theta)$$



LONDON. Um eine Operation für seine erkrankte kleine Tochter zahlen zu können, bot ein Brite seine Niere im Internet zur Versteigerung an. Der verzweifelte Vater offeriert sein Organ im Rahmen einer Online-Auktion. Er hoffte auf einen Erlös von 100.000 Pfund, wie eine englische Zeitung berichtete.

Blickbewegungsmodelle

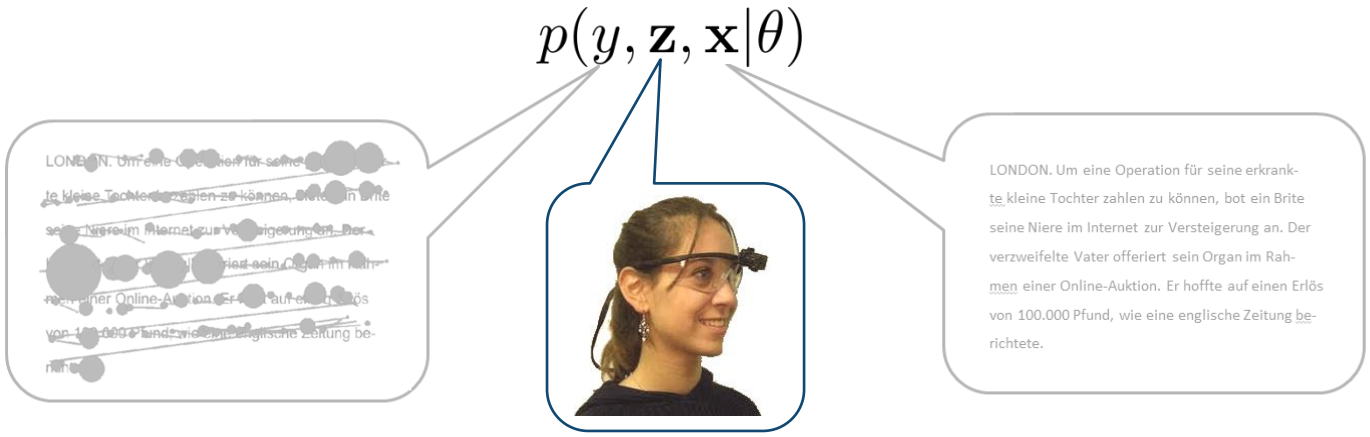
- **Ziel**: gemeinsame Modelle von Blickbewegungen, Textinhalt und Attributen des Lesers.



- Vorhersage der Blickbewegung (Psychologie).

Blickbewegungsmodelle

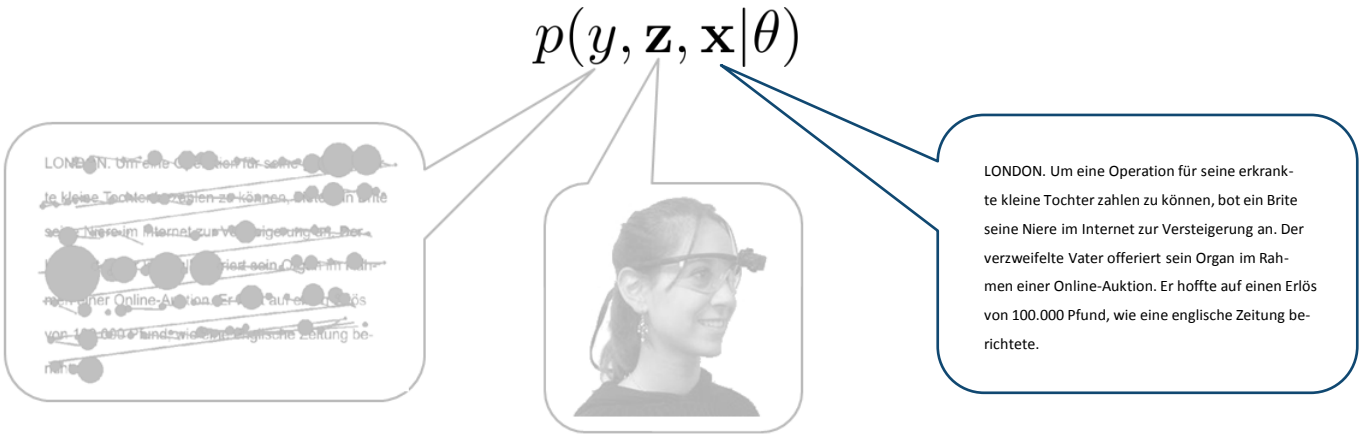
- **Ziel**: gemeinsame Modelle von Blickbewegungen, Textinhalt und Attributen des Lesers.



- Rückschlüsse auf die Identität des Lesers (Biometrie).
- Inferenz von Eigenschaften des Lesers: Muttersprache, Alter, Kompetenz (Textpersonalisierung, Kompetenztests).

Blickbewegungsmodelle

- **Ziel**: gemeinsame Modelle von Blickbewegungen, Textinhalt und Attributen des Lesers.



- Rückschluss auf die Lesbarkeit von Texten, durch Vorhersage der zu erwartenden Blickbewegungen (Werkzeuge zur Textanalyse).

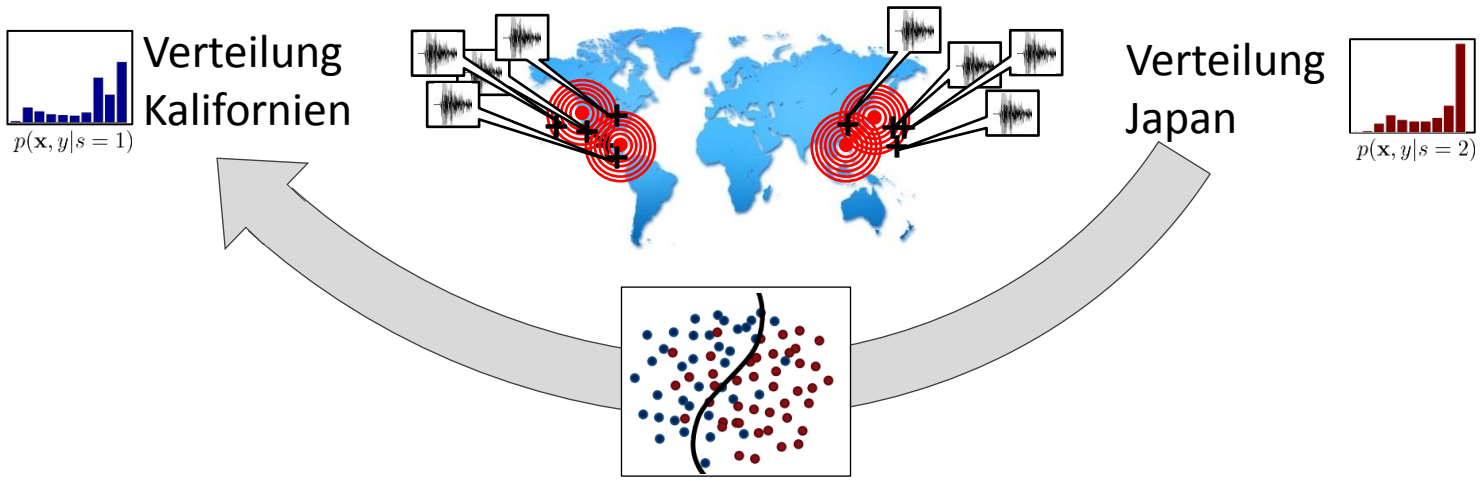
Übersicht

- Einleitung & Motivation
- Lernen unter Datenabhängigkeiten
- Lernen unter Verteilungsverschiebungen
- Modellevaluierung

Transferlernen

- Messungen an unterschiedlichen Orten: Verteilungsverschiebungen.
- Standardansatz: Verteilungsunterschiede ignorieren und Daten poolen.
- **Transferlernen**: Verteilungsunterschiede zwischen Teilmengen der Daten korrigieren.

Transferlernen: Vorstudie

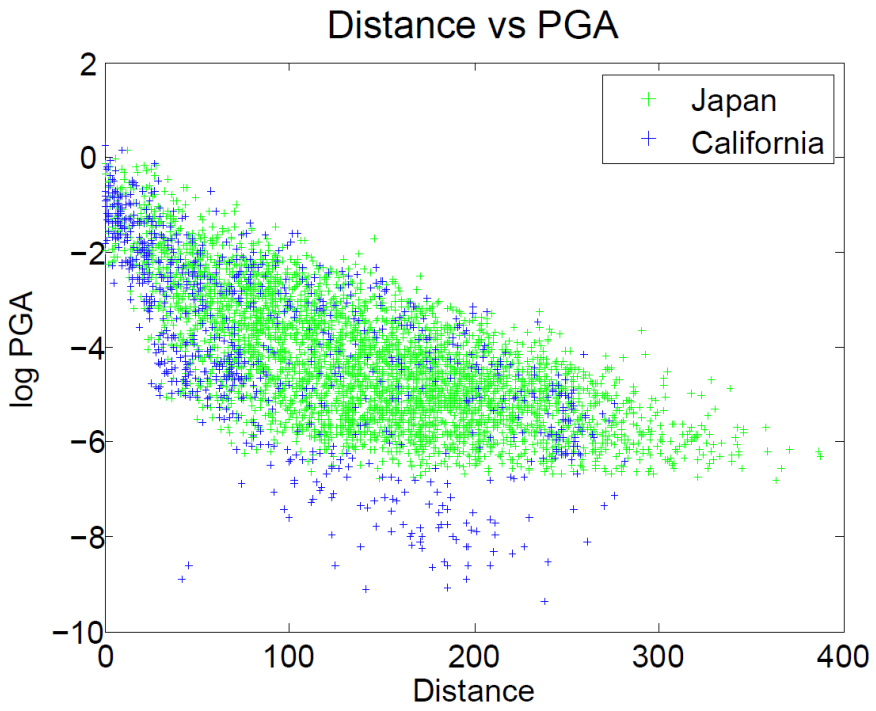


- Vorstudie zum Transferlernen, mit AG Frank Scherbaum, Geophysik.
- Ansatz der Verteilungskorrektur [Bickel, Sawade, Scheffer; 2008]
  - Annahme: jeweils einheitliche Verteilung für Daten aus Kalifornien, Japan.
  - Explizites Modell des Verteilungsunterschieds aus Daten lernen.

Erdbebendaten Japan/Kalifornien

- Bodenbeschleunigungsdaten aus Kalifornien (841) und aus Japan (3542)
- Ziel: Modell für Kalifornien („Zielverteilung“/“Zieldaten“)
- Zusätzlich können Daten aus Japan verwendet werden („Hilfsdaten“)

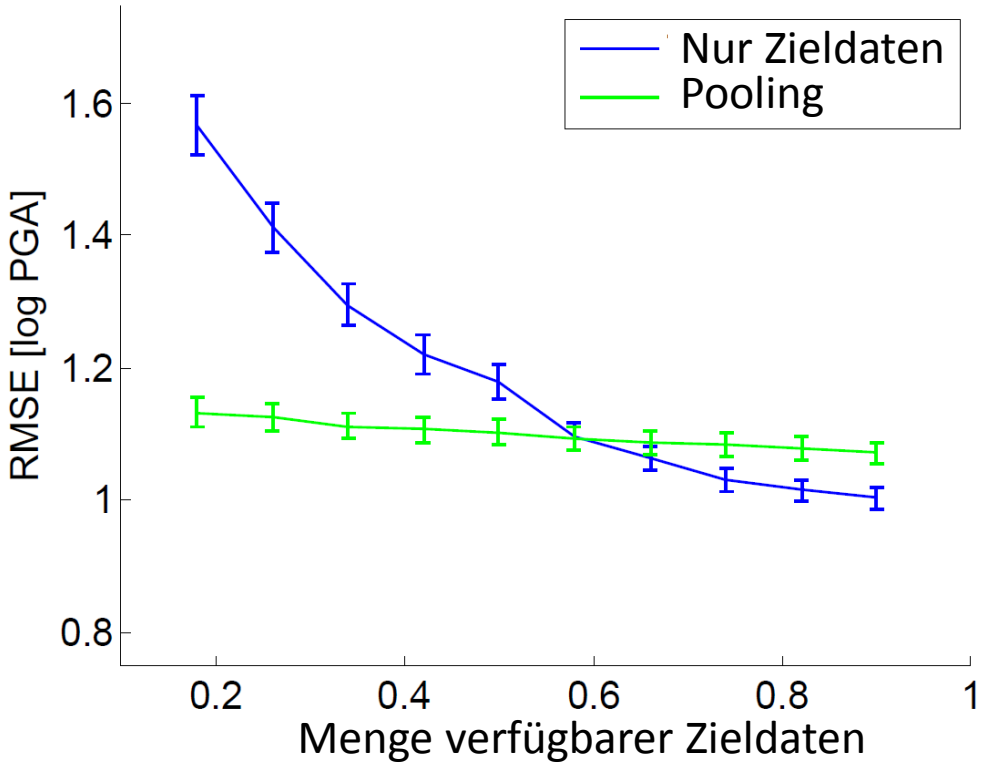
Leicht unterschiedliche Verteilung über Daten aus Kalifornien und Japan



Vergleich “Nur Zieldaten”, “Pooling”

- Baseline 1: “**Nur Zieldaten**” (ignoriert Hilfsdaten)
- Baseline 2: “**Pooling**” (ignoriert Verteilungsunterschied)

Fehler in Abhängigkeit der zum Training verfügbaren Zieldaten: Bias vs Varianz



Verteilungskorrektur

- Transferlernen (Verteilungskorrektur)

  Zieldaten (Kalifornien): $(\mathbf{x}, y) \sim p(\mathbf{x}, y \mid t = 1)$

  Hilfsdaten (Japan): $(\mathbf{x}, y) \sim p(\mathbf{x}, y \mid t = 0)$

- Idee: Umgewichten der Daten, so dass $p(\mathbf{x}, y \mid t = 0)$ zu $p(\mathbf{x}, y \mid t = 1)$ korrigiert wird.

Verteilungskorrektur

- Verteilungskorrektur formal
  - Pool folgt Mischverteilung

$$p(\mathbf{x}, y) = \underbrace{p(\mathbf{x}, y \mid t=1)}_{\text{Zieldaten}} \underbrace{p(t=1)}_{\text{Anteil Zieldaten}} + \underbrace{p(\mathbf{x}, y \mid t=0)}_{\text{Hilfsdaten}} \underbrace{p(t=0)}_{\text{Anteil Hilfsdaten}}$$

  - Wir hätten gerne alle Daten aus Zielverteilung $p(\mathbf{x}, y \mid t=1)$
  - Daten umgewichten mit $\mathrm{r}(\mathbf{x}, y) = \dfrac{p(\mathbf{x}, y \mid t=1)}{p(\mathbf{x}, y)}$

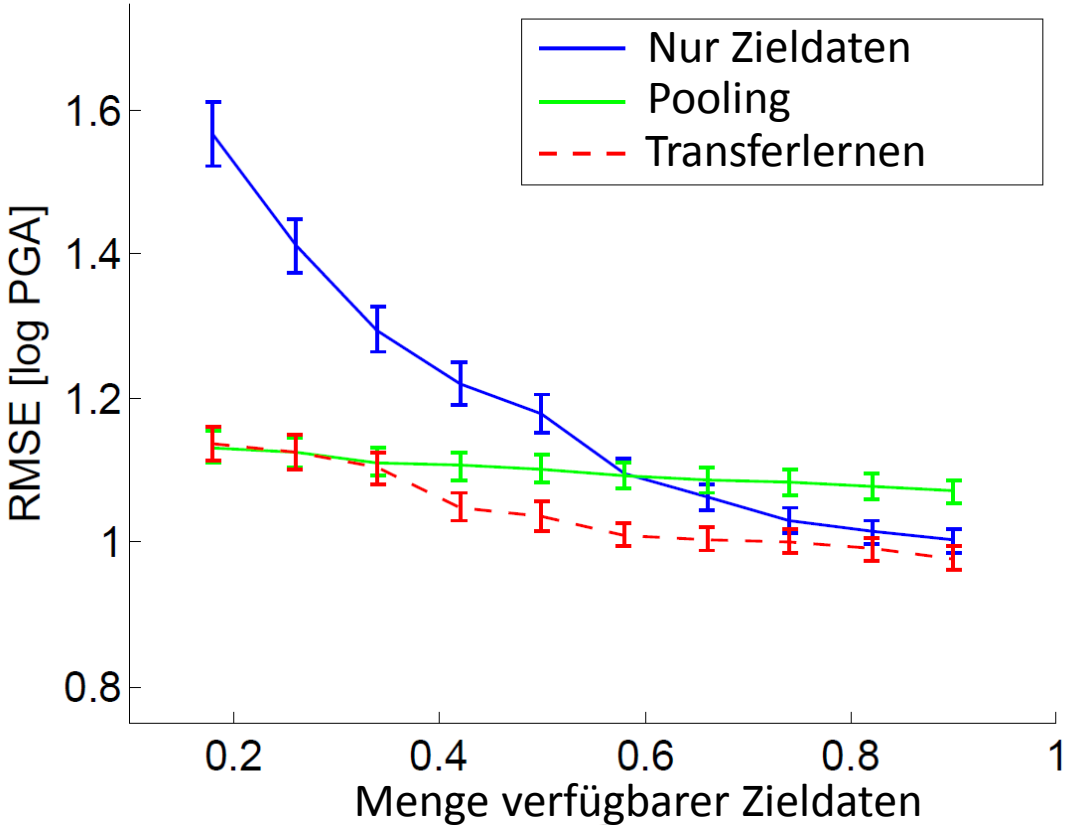
- Umgewichtungsfaktoren mit Modell schätzen

$$\mathrm{r}(\mathbf{x}, y) = \frac{p(\mathbf{x}, y \mid t=1)}{p(\mathbf{x}, y)} = \frac{p(t=1 \mid \mathbf{x}, y)\, p(\mathbf{x}, y)}{p(\mathbf{x}, y)\, p(t=1)} \propto p(t=1 \mid \mathbf{x}, y)$$

Diskriminatives Modell: Zieldaten gegen Hilfsdaten (hier: logistische Regression, Poly-Kernel)

Ergebnisse: Testfehler

- Vergleich “**Nur Zieldaten**”, “**Pooling**”, und “**Transferlernen**”

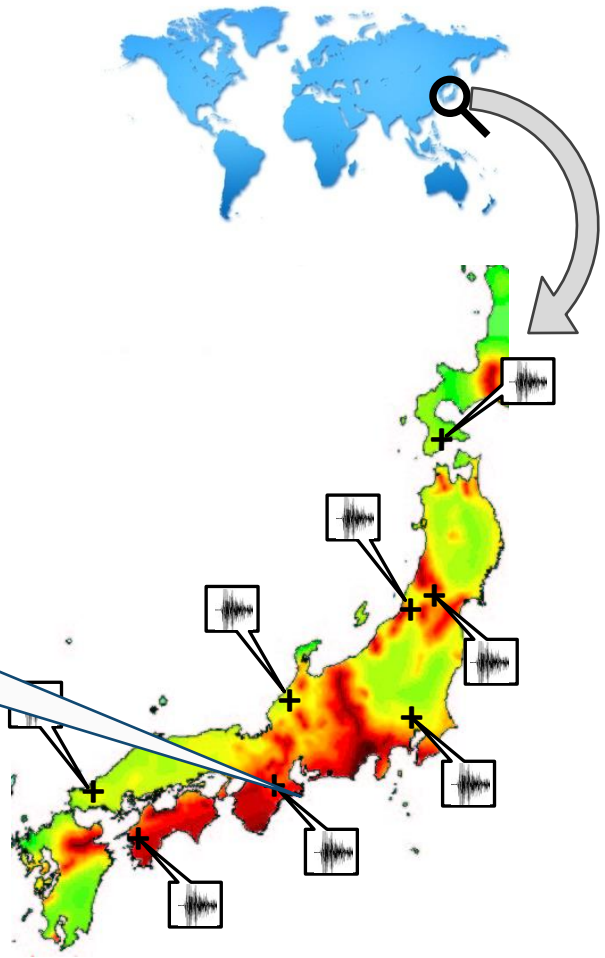


Transferlernen: Ziele

- Aufteilung in Teilmengen Vereinfachung: Verteilung ändert sich kontinuierlich mit dem Ort der Messung.

- **Ziel**: Modelle, die eine kontinuierliche Verteilungsänderung abbilden.

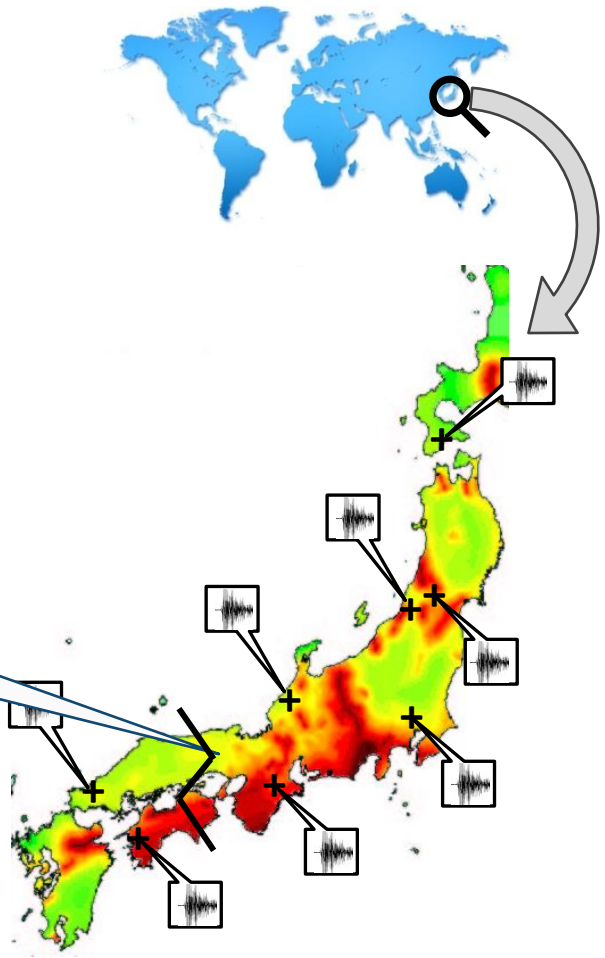
Vorhersage für Standort:
Aufwendige Inferenzprobleme,
approximative Inferenzalgorithmen.



Transferlernen: Ziele

- Aufteilung in Teilmengen Vereinfachung: Verteilung ändert sich kontinuierlich mit dem Ort der Messung.
- **Ziel**: Modelle, die geophysikalisches Vorwissen berücksichtigen.

Bekannte seismische Bruchlinie:
stärkere Verteilungsänderung erwartet.

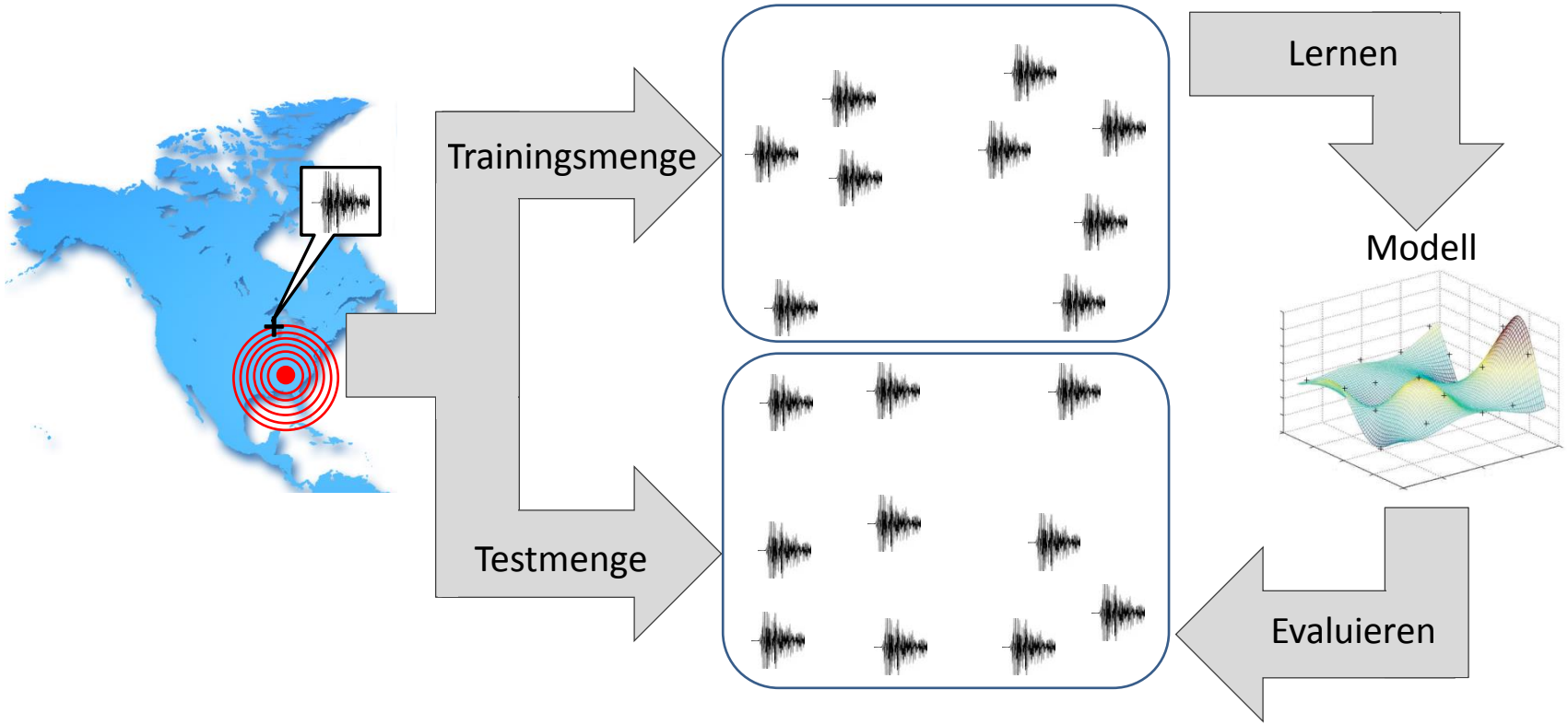


Übersicht

- Einleitung & Motivation
- Lernen unter Datenabhängigkeiten
- Lernen unter Verteilungsverschiebungen
- Modellevaluierung

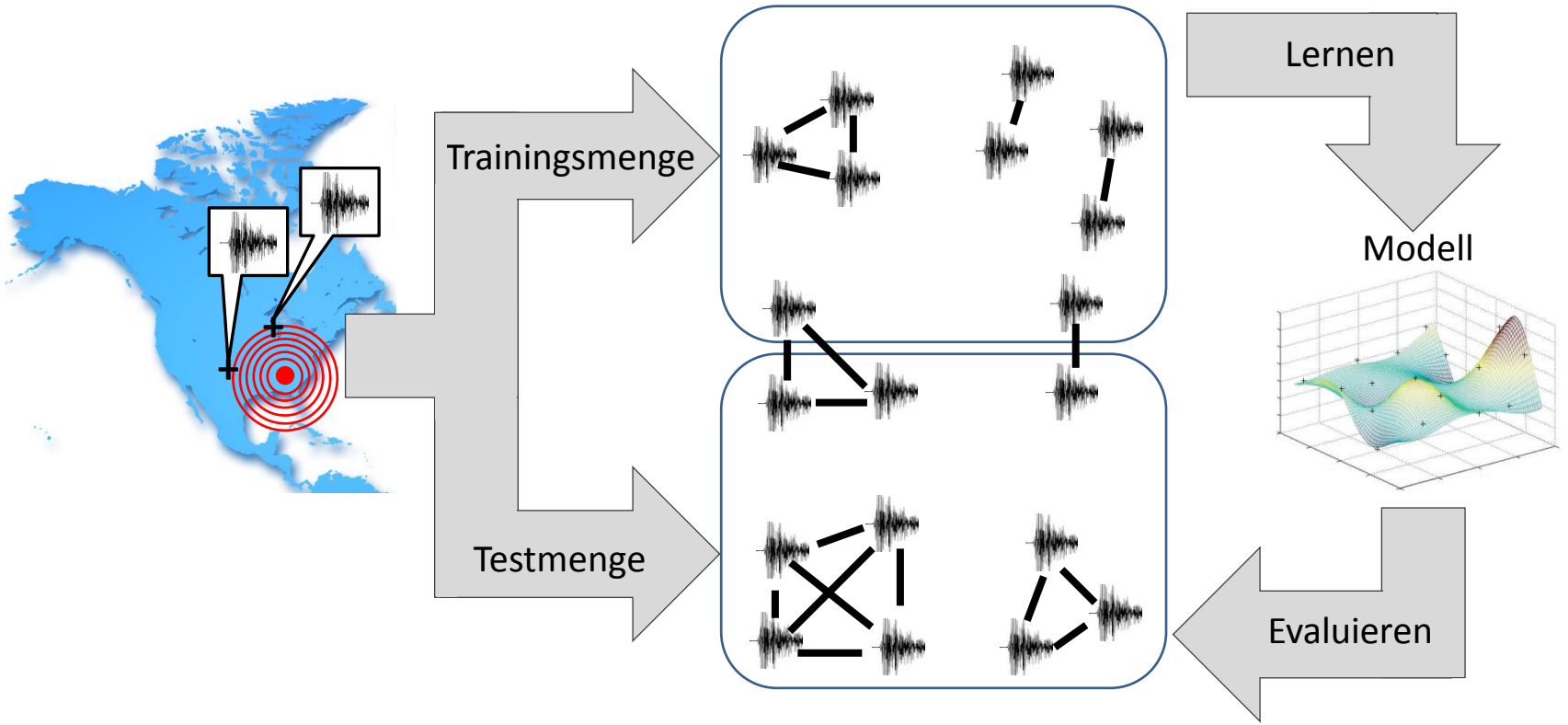
Evaluierung: Abhängigkeiten

- Existierende Evaluierungsverfahren basieren auf der Annahme unabhängiger und einheitlich verteilter Beobachtungen.



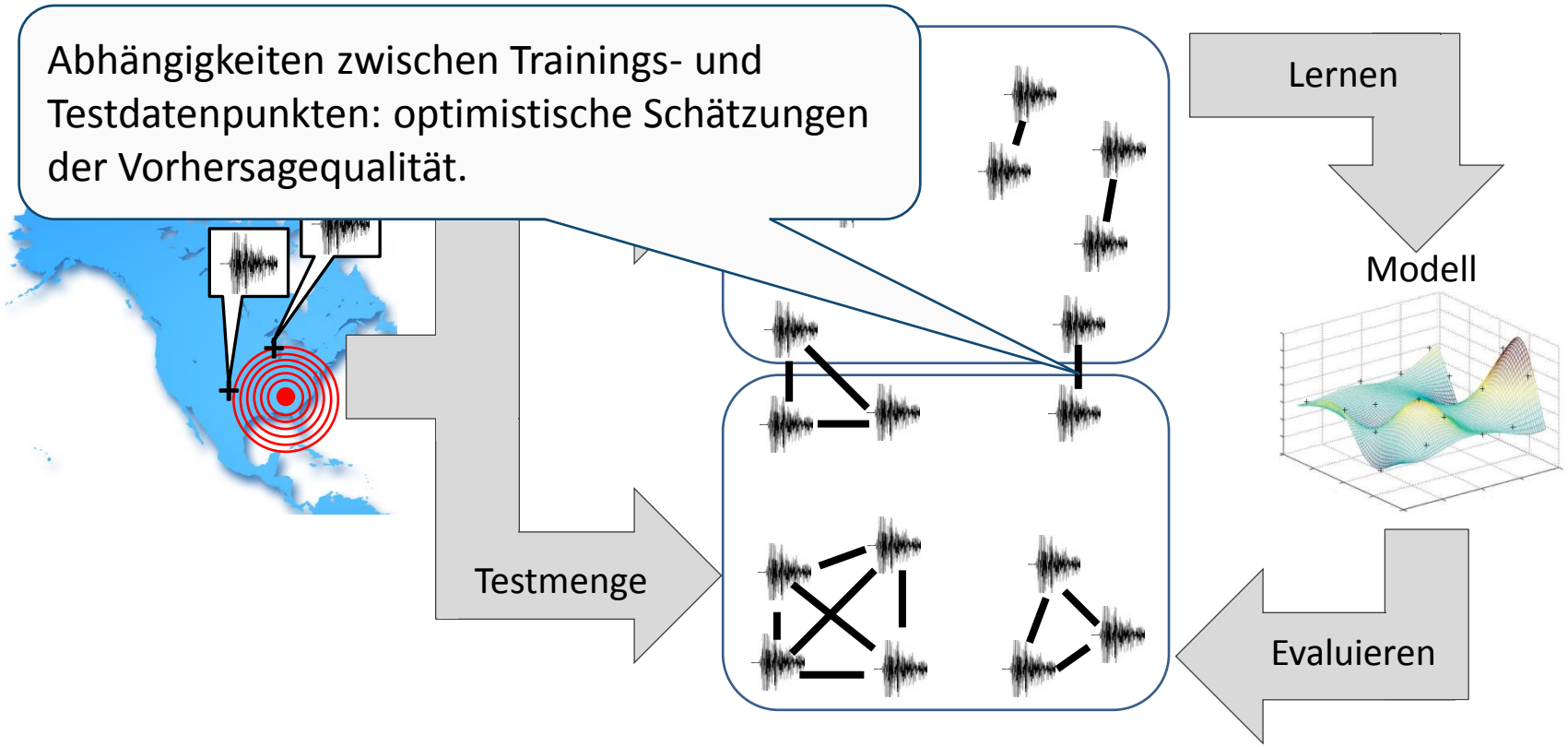
Evaluierung: Abhängigkeiten

- Abhängige Datenpunkte verfälschen das Ergebnis bekannter Evaluierungs-verfahren in verschiedener Weise.



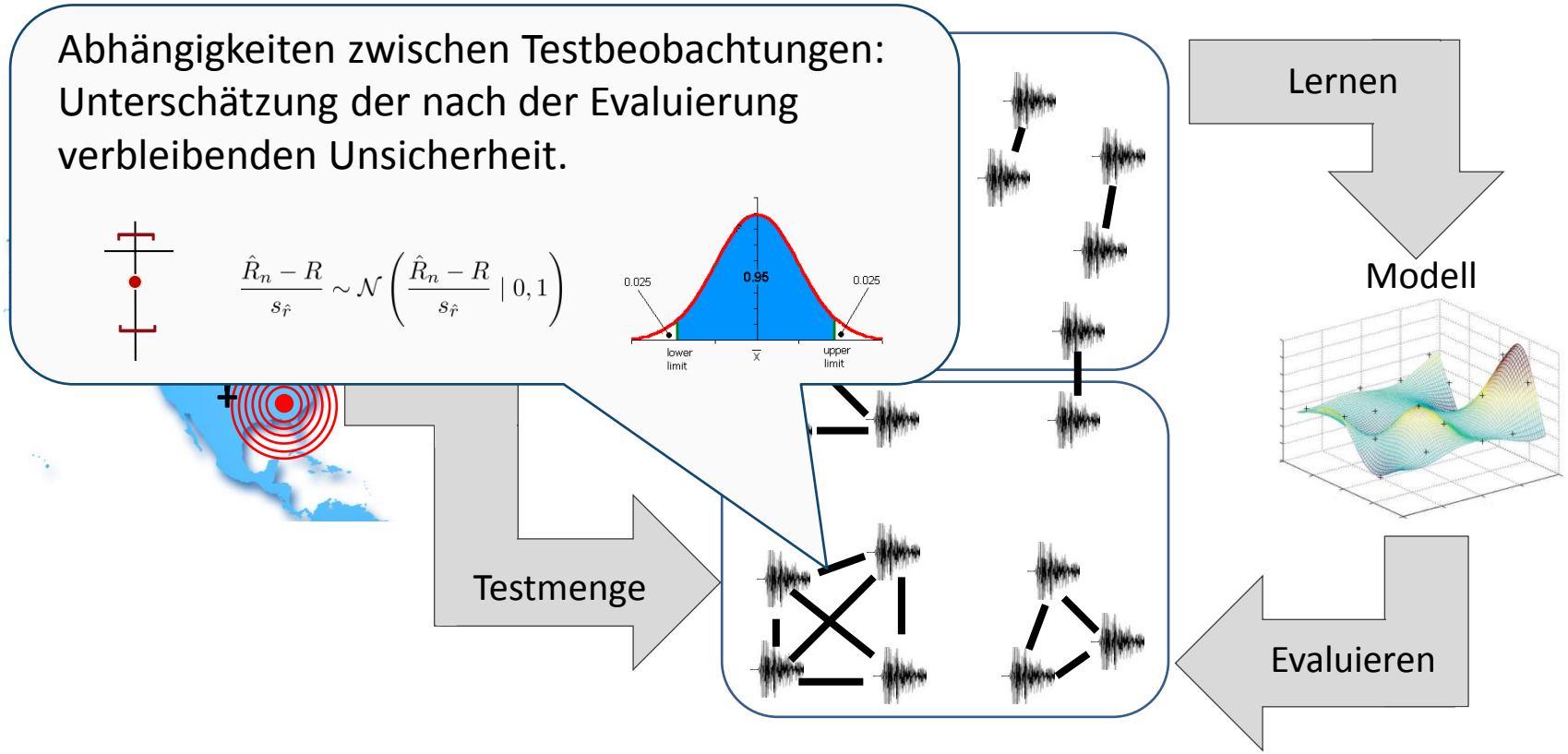
Evaluierung: Abhängigkeiten

- Abhängige Datenpunkte verfälschen das Ergebnis bekannter Evaluierungs-verfahren in verschiedener Weise.



Evaluierung: Abhängigkeiten

- Abhängige Datenpunkte verfälschen das Ergebnis bekannter Evaluierungs-verfahren in verschiedener Weise.



Evaluierung: Verteilungsverschiebungen

Blickbewegungen: Studenten

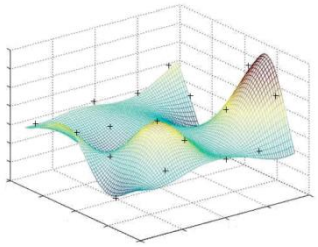
Beobachtung

Daten

ID	D	F	M	R	I	H	PGA
1	100	305	4,50	2,12	0	3,15	2,12
2	122	400	3,10	1,75	0	2,05	1,75
3	231	265	5,40	2,45	1	3,54	2,45
4	146	300	3,80	1,93	1	1,32	1,93
5	133	195	4,90	2,44	1	2,44	2,44
6	215	245	3,20	1,63	0	1,33	1,63
7	100	305	4,50	2,12	1	1,33	2,12
8	122	400	3,10	1,75	0	1,98	1,75
9	231	265	5,40	2,45	1	2,05	2,45
10	146	300	3,80	1,93	1	3,54	1,93
11	133	195	4,90	2,44	1	1,32	2,44
12	215	245	3,20	1,63	0	2,44	1,63
13	251	310	6,50	2,83	1	1,33	2,83
14	133	195	4,90	2,44	1	1,33	2,44
15	215	245	3,20	1,63	1	3,21	1,63

Modellbildung

Modell



Beschreibung, Vorhersage

Evaluierung: Verteilungsverschiebungen

Wie gut beschreibt das mit studentischen Probanden gelernte Modell die Blickbewegungen der Gesamtbevölkerung?

Blickbewegungen: Studenten

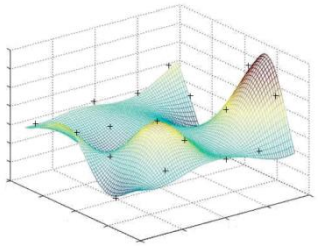
Beobachtung

Daten

ID	D	F	M	R	I	H	PGA
1	100	305	4,50	2,12	0	3,15	2,12
2	122	400	3,10	1,75	0	2,05	1,75
3	231	265	5,40	2,45	1	3,54	2,45
4	146	300	3,80	1,93	1	1,32	1,93
5	133	195	4,90	2,44	1	2,44	2,44
6	215	245	3,20	1,63	0	1,33	1,63
7	100	305	4,50	2,12	1	1,33	2,12
8	122	400	3,10	1,75	0	1,98	1,75
9	231	265	5,40	2,45	1	2,05	2,45
10	146	300	3,80	1,93	1	3,54	1,93
11	133	195	4,90	2,44	1	1,32	2,44
12	215	245	3,20	1,63	0	2,44	1,63
13	251	310	6,50	2,83	1	1,33	2,83
14	133	195	4,90	2,44	1	1,33	2,44
15	215	245	3,20	1,63	1	3,21	1,63

Modellbildung

Modell



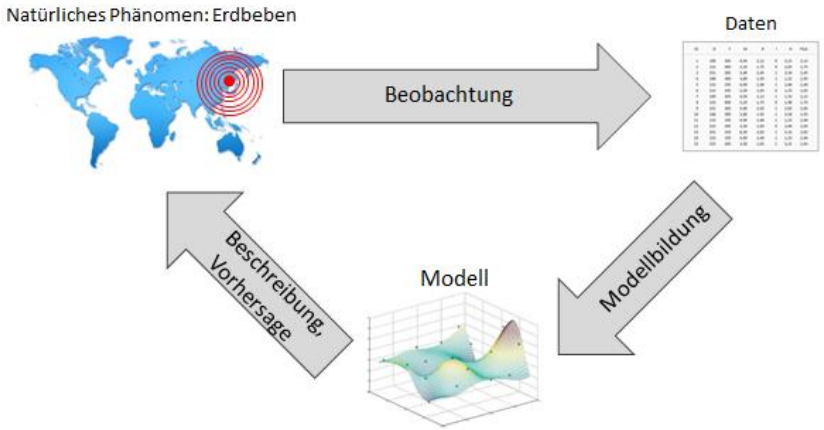
Beschreibung, Vorhersage

Blickbewegungen: Gesamtbevölkerung

Übersicht

- Einleitung & Motivation
- Lernen unter Datenabhängigkeiten
- Lernen unter Verteilungsverschiebungen
- Modellevaluierung

Zusammenfassung



- Maschinelles Lernen zur Modellbildung aus Beobachtungsdaten in den Naturwissenschaften.
- Besondere Verteilungseigenschaften, insbesondere **Abhängigkeiten** und **Verteilungsverschiebungen**.
- Zusammenarbeit mit Arbeitsgruppen aus der Geophysik und der kognitiven Psychologie.