

Teile und (Be-) Herrsche

Einführung in Bayessche Hierarchische Modelle

IFM Workshop, 28.01.2021

Dr. Sven Thies

AGENDA

1. Theorie

- i. Machine Learning vs. Statistische Modellierung
- ii. Grundlagen der Bayesschen Modellierung
- iii. Pymc3 – Ein probabilistisches Programmierframework

2. Praxis

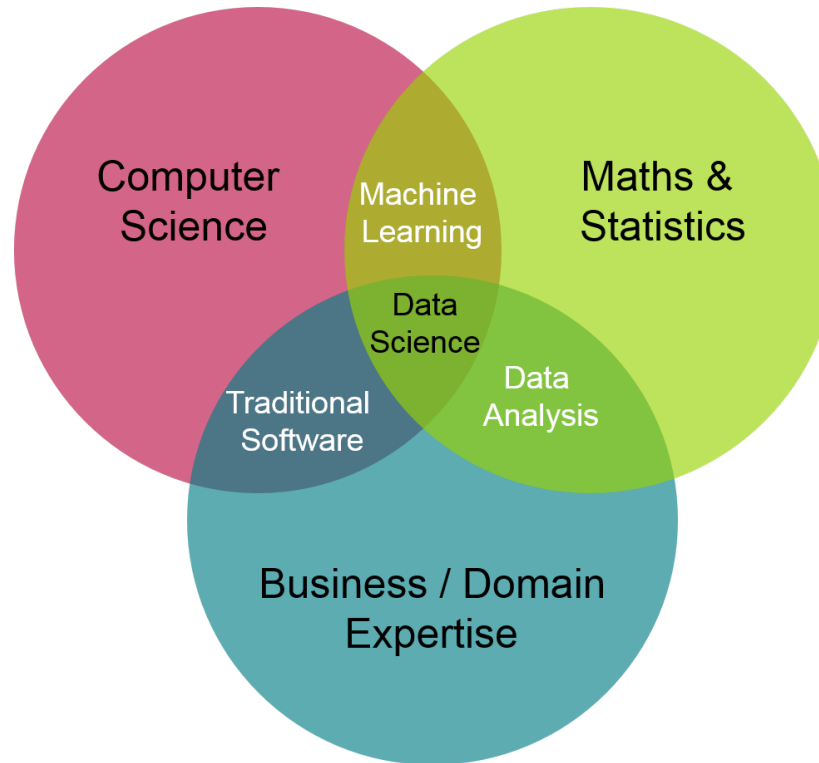
- i. Unsicherheit beherrschen: Bayessche Lineare Modelle
- ii. Informationen teilen: Bayessche Hierarchische Modelle

Machine Learning vs. Statistische Modellierung

Machine Learning vs. Statistische Modellierung

Unterschiedliche Hintergründe bestimmen Weltbild und Herangehensweise

Der Informatiker



Der Statistiker



Machine Learning vs. Statistische Modellierung

Data Science Stereotypen

... als Data Scientist

Der Informatiker



- Verwendet Python (TF, sklearn)
- Große Datensätze sind notwendig
- Sorgt sich um Overfitting
- Prüft die Vorhersagekraft
- Verwendet nichtlineare Modelle
- Modelle sind oft Black-Box
- Strebt nach Automatisierung**

Der Statistiker



Quelle: In Anlehnung an „Machine Learning and Statistics: Don't Mind the Gap“, Thomas Wiecki, ODSC Europe, 2018

Machine Learning vs. Statistische Modellierung

Herangehensweise bei der Modellierung

... als Data Scientist

Der Informatiker



Viele Daten
Große Modelle

Regularisierung

Finales Modell

➔ **Top-Down**

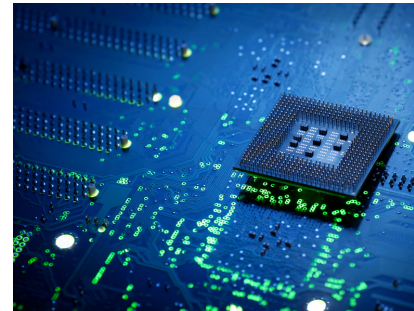
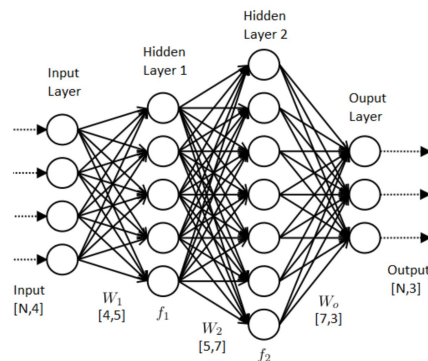
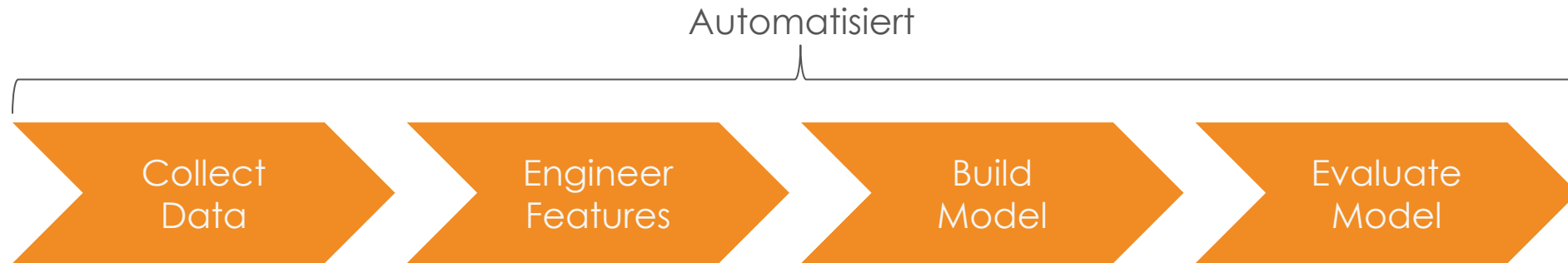
Machine Learning-Ansatz

Der Statistiker



Machine Learning vs. Statistische Modellierung

Prozess der Modellierung nach Informatikern



Machine Learning vs. Statistische Modellierung

Data Science Stereotypen

... als Data Scientist

Der Informatiker



Verwendet Python (TF, sklearn)

Große Datensätze sind notwendig

Sorgt sich um Overfitting

Prüft die Vorhersagekraft

Verwendet nichtlineare Modelle

Modelle sind oft Black-Box

Strebt nach Automatisierung

Der Statistiker



Verwendet R

Große Datensätze sind Problem

Sorgt sich um Annahmen

Prüft asymptotisches Verhalten

Verwendet lineare Modelle

Modelle sind verständlich

Strebt nach Erkenntnis

Quelle: In Anlehnung an „Machine Learning and Statistics: Don't Mind the Gap“, Thomas Wiecki, ODSC Europe, 2018

Machine Learning vs. Statistische Modellierung

Herangehensweise bei der Modellierung

... als Data Scientist

Der Informatiker



Viele Daten
Große Modelle

Regularisierung

Finales Modell

➔ **Top-Down**

Machine Learning-Ansatz

Finales Modell

Experimente

Wenige Daten
Kleine Modelle

Bottom-Up ➔

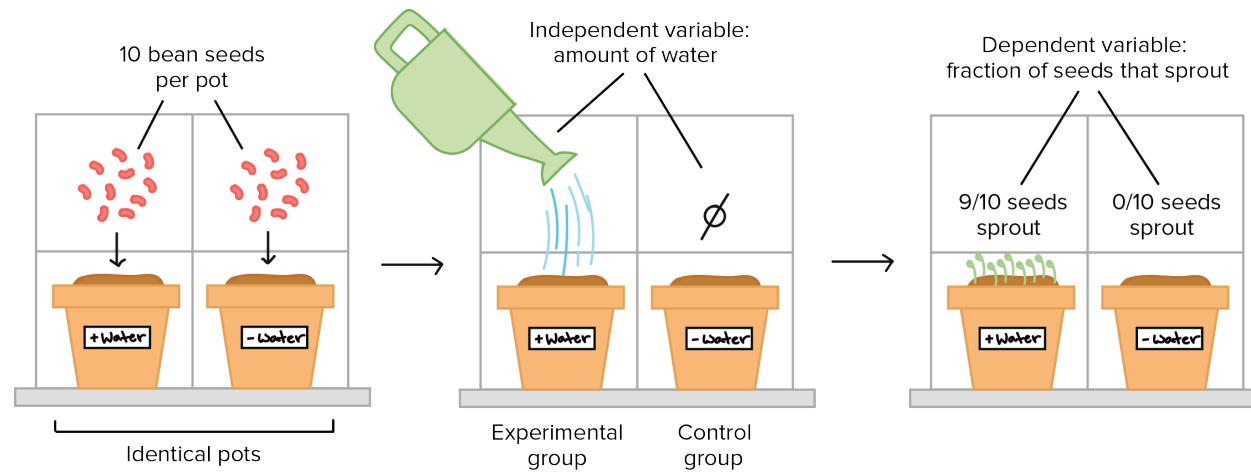
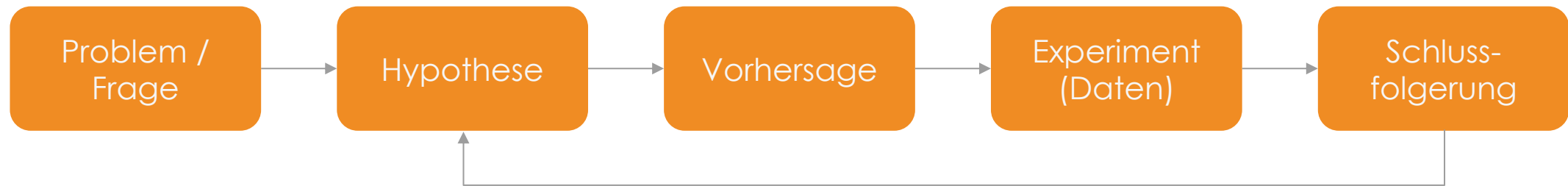
Inferenz-basierter Ansatz

Der Statistiker



Machine Learning vs. Statistische Modellierung

Prozess der Modellierung nach Statistikern



Grundlagen der Bayesschen Modellierung

Grundlagen der Bayesschen Modellierung

Mathematische Formulierung des wissenschaftlichen Prozesses: Der Satz von Bayes



1701 – 1761

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)}$$

$$P(H|D) \propto P(D|H)P(H)$$



Grundlagen der Bayesschen Modellierung

Das Posteriori Wissen ist eine Kombination von A-Priori Wissen und Beobachtungen

Posterior Probability

Die Wahrscheinlichkeit, dass die Hypothese wahr ist, nachdem wir die Daten gesehen haben.

Likelihood

Die Wahrscheinlichkeit der beobachteten Daten, angenommen, die Hypothese sei wahr.

Prior Probability

Die Wahrscheinlichkeit, dass die Hypothese Wahr ist (bevor wir die Daten gesehen haben).

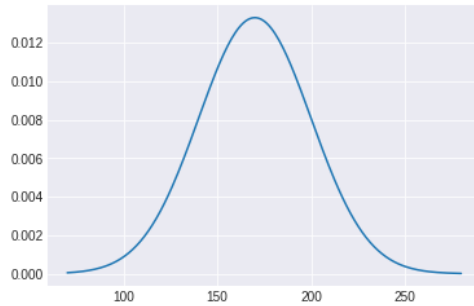
$$P(H|D) \propto P(D|H)P(H)$$

Alles wird als Verteilung dargestellt

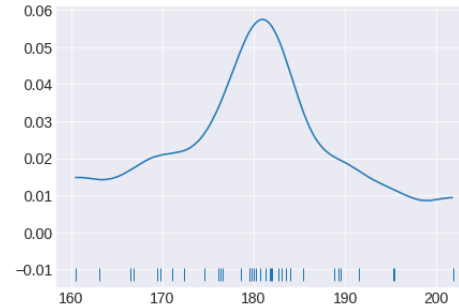
Vorwissen wird genutzt

Grundlagen der Bayesschen Modellierung

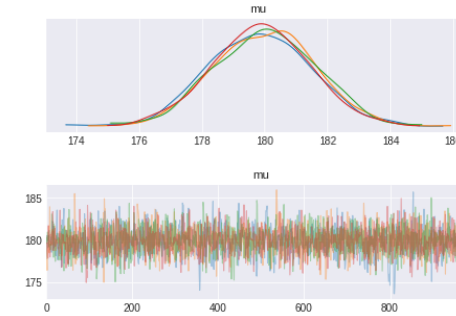
Prozess der Bayesschen Modellierung



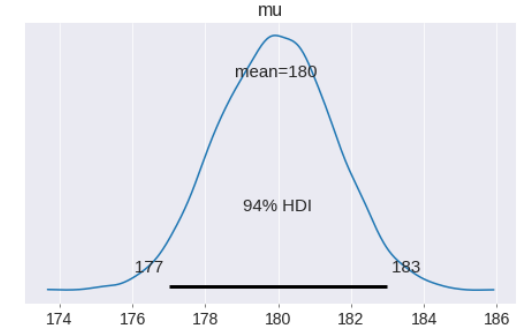
$$P(\mu) \rightarrow \mu \sim N(170, 30)$$



$$P(y|\mu) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}$$



$$P(\mu|y) \propto P(y|\mu, \sigma)P(\mu)$$



Bayesian Updating (of believes)

Pymc3 – Ein probabilistisches Programmierframework

Pymc3 – Ein probabilistisches Programmierframework

Vereinfachte Bayessche Modellierung mit Pymc3

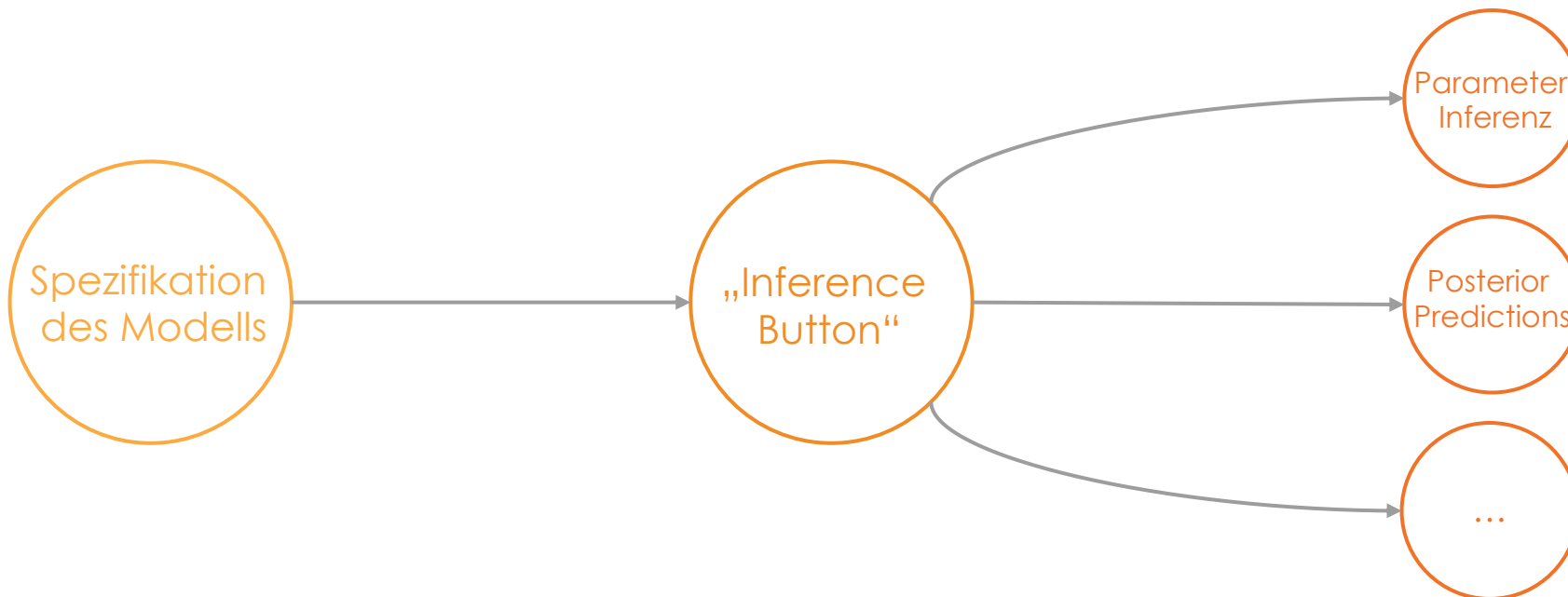


- Leicht zu bedienende API
- Starkes Backend für Sampling

1. Schritt

2. Schritt

3. Schritt



Praxis