# Folie 1

* Atanas Dimitrov
* Automatische Auswahl von maschinellen Lernverfahren für kausale Inferenz

# Folie 2 Blitzvorstellung

* Synth-Validation: Existierende Methode zur automatischen Auswahl implementiert
* R
* Methoden aus dem maschinellen Lernen
* Auf unterschiedlichen Datensätzen getestet – echte Rohdaten, synthetische Daten, zufällig generierte Daten

# Folie 3 Überblick

* Kausale Inferenz
  + Was?
  + Fundamentales Problem
  + Hauptaufgabe
  + Problematik
* Synth-Validation
  + Warum ist eine (automatische) Auswahl notwendig?
  + Funktionsweise, benutzte Algorithmen und Heuristiken
* Evaluation
  + Art und Herkunft der Daten
  + Vorgehensweise beim Benchmarking
  + Vorstellung der Ergebnissen

# Folie 4 Kausale Inferenz

* Ist nicht das Suchen nach der Ursache
* Sondern eine potenzielle Ursache ist bekannt
* Beitrag, Ergebnis hängt auch von weiteren Faktoren ab
* Rauchen -> Gesundheit
* Effekt – Differenz zwischen mit und ohne Behandlung

# Folie 5 Das fundamentale Problem

* Das Problem
* Differenz zwischen den beiden Zuständen in nicht direkt messbar

# Folie 6 Schätzen von dem durchschnittlichen Behandlungseffekt

* Schätzung für den durchschnittlichen Behandlungseffekt einer Grundgesamtheit
* Differenz zwischen den Erwartungswerten der potenziellen Ergebnissen – behandelt und unbehandelt

# Folie 7 Datensammlung

* Experiment
  + Zufällig unbehandelte Objekte auswählen
  + Zwei Gruppen bilden
  + Behandlungsgruppe behandeln
  + Differenz der Erwartungswerte
  + Gute Schätzung, frei von Bias
  + T-test durchführen, Konfidenzintervalle bilden
  + Nicht immer möglich – Zeitspanne der Behandlung, unethische Behandlung
* Beobachtungsdaten
  + Im Allgemein eine verzerrte Schätzung mit der naiven Methode
  + Spezielle Methoden werden benutzt

# Folie 8 Confounding

* Grafik beschreiben

# Folie 9 Methoden für kausale Inferenz

* Ziel: Confounding und Bias vermindern/entfernen
* Gruppen anpassen mit Propensity Score
* Keine allgemeinbeste Methode

# Folie 10 Synth-Validation

* Die beste Methode für einen Datensatz auswählen
* Generierung von synthetischen Daten mit vorgegebenen Effekten, ähnlich
* Ausführung von den Methoden auf diesen Daten
* Auswahl von der Methode mit dem kleinsten durchschnittlichen Fehler

# Folie 11 Synth-Validation

* Abbildung beschreiben
* Synthetische Effekten berechnen, Schätzungen aus den Methoden, Mediane, Spannweite
* Constrained Boosting, bedingten Erwartungswerten für jedes x anpassen
* Stichprobe aus (x,w) ziehen, sie in Mues einsetzen, Daten generieren
* Den Effekt von diesen Daten schätzen

# Folie 12 Daten

* Unterschiedliche Datenarten – echten, synthetischen und zufällig generieten.
* Nur die allgemeinen Ergebnissen vorstellen
* Mit einem bekannten echten Behandlungseffekt sein
* Aus Wettbewerben für kausale Inferenz
* Nicht vorgegeben, was hinter die Behandlungsvariable und die Ergebnisvariable steht

# Folie 13 Benchmarking

* Integrierte Methoden
* Jeder Datensatz mindestens 30 mal gebenchmarkt
* Insgesamt 373 Läufe
* Einen Lauf sehr aufwendig – durchschnittlich mehr eine Stunde

# Folie 14 Evaluation

* Die Ergebnisse vorstellen
* Barplot mit dem MAE von den Schätzungen von allen Methoden
* MAE – absolute Fehler zwischen dem echten Behandlungseffekt und der Schätzung
* Orakel an erster Stelle, wählt immer die beste Schätzung aus, keine echte Methode
* Ziemlich hohe Fehler, hohe Varianz in den Schätzungen
* R\_boost mit dem kleinsten durchschnittlichen Fehler, erst dann Synth-Validation
* Gegen unsere Hypothese, dass Synth-Validation der beste Schätzer ist
* Synth-Validation an der zweiten Stelle, Ergebnis nicht signifikant, Welch Test
* Erstaunlich, dass das Lasso schlechter schätzt als die einfachste Methode
* Lässt sich zu einigen heftigen Ausreißern zurückzuführen
* Mediane von dem Lasso niedriger

# Folie 15 Evaluation

* Balkendiagramm mit der Auswahlhäufigkeit von dem Orakel und Synth-Validation
* Wie oft jede getestete Methode von dem Orakel und Synth-Validation ausgewählt ist
* Blaue Balke – Orakel, lila Balke – Synth-Validation
* Jede Methode ist irgendwann mal die beste
* R\_boost am häufigsten die beste Schätzung
* Kaum von Synth-Validation ausgewählt
* Wenn Synth-Validation gut ausgewählt hätte, sollten jede zwei Balken überreinstimmen

# Folie 16 Evaluation

* Tabelle mit der kumulierter Erfolgsrate von Synth-Validation
* Zeigt wie oft Synth-Validation die beste Schätzung, die zweitbeste Schätzung usw. ausgewählt hat
* Dritte und vierte Spalte
* Die interessanteste ist die Erfolgsrate für die beste Schätzung
* Etwa höher als die zufällige Auswahl von einer Methode
* Nicht genug, um signifikant zu sein
* Unter den ersten zwei oder drei – signifikant besser als der Zufall

# Folie 17 Fragen