

Datenaugmentation in klinischen Studien

M.Sc. Marie Diehl

marie.diehl@uni-marburg.de

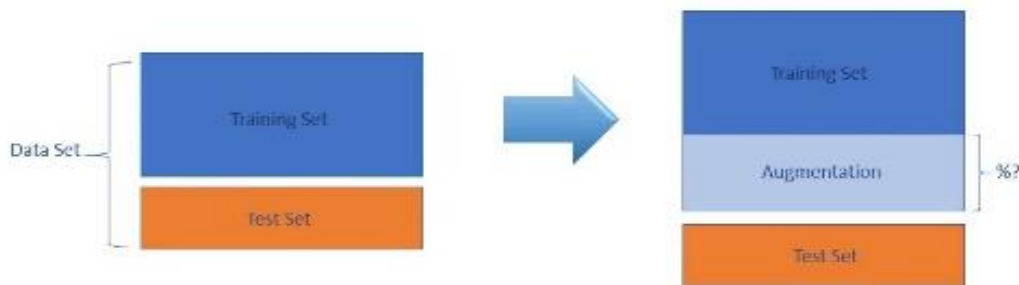
(Prof. Dr. Dominik Heider)

Zusammenfassung

Bei der Auswertung und Analyse klinischer Studien mit Hilfe von machine learning, ist die Datenmenge oft stark beschränkt. Je geringer der Datenumfang, umso schwerer gestaltet sich eine aussagekräftige Vorhersage mit machine learning Methoden. Mit Hilfe von Datenaugmentation sollen hier die Klassifizierung von Daten verbessert werden.

Eine Möglichkeit der Datenaugmentation besteht darin, per Zufall (bootstrapping) bestimmte Datenpunkte zu vervielfältigen und diese dann mit einem zufälligen Noise zu verrauschen. Darüber hinaus gibt es die LDA-Methode [1,3,4] oder einen Bayesischen Schätzer [2] welche zur Datenaugmentation bei geringer Datenmenge verwendet werden.

In dieser Arbeit soll insbesondere der Umfang dieser Datenaugmentation untersucht werden und in welchem Umfang sich diese positiv oder negativ auf die erfolgreiche Anwendung des machine learnings auswirkt.



Literatur

- [1] Sharma A, Paliwal K: Linear discriminant analysis for small size problem: an overview. Int.J.Mach.Learn.&Cyber. (2015) 6:433-454
- [2] McNeish D: On Using Bayesian Methods to Adress Small Sample Problems. Structural Equation Modeling:A Multidisciplinary Journal (2016), 23:2, 750-773
- [3] Liu J, Chen S et al.: A study on three linear discriminant analysis based methods in small sample size problem. Pattern Recognition (2008) 41,102-226
- [4] Tebbens J, Schlesinger P: Improving implementation of linear discriminant analysis for the high dimension/small sample size problem. Computational Statistics & Data Analysis (2007) 52,423-437

Bitte melden Sie sich bei Interesse per Email bei:

M.Sc.Marie Diehl marie.diehl@uni-marburg.de

Stand: 01.02.2020