

Projektion von Machine Learning Scores auf Wahrscheinlichkeiten

Prof. Dr. Dominik Heider
dominik.heider@uni-marburg.de

Zusammenfassung

Statistische und maschinelle Lernverfahren werden bereits in der medizinischen Diagnostik und Prognostik eingesetzt, wie z.B. bei Lebererkrankungen [1], HIV [2] oder zur Klassifikation von Gehirndaten [3]. Dabei werden in der Regel überwachte Lernverfahren eingesetzt, welche auf einem vorhandenen Datensatz trainiert und dann mit weiteren Daten unabhängig evaluiert werden. Für diese neuen Daten wird dann ein Score [4] berechnet und anhand eines cutoffs dann die zugehörige Klasse identifiziert. Zur Interpretation der Ergebnisse wäre es jedoch besser nicht nur einen Score bzw. eine Klassenzuordnung zu erhalten, sondern zudem auch eine Wahrscheinlichkeit für die Zugehörigkeit zu einer Klasse. Diese Wahrscheinlichkeiten werden bei der Bayes Klassifikation bereitgestellt, jedoch bieten andere Verfahren, wie z.B. support-vector Maschinen, neuronale Netze oder Random Forests, diese Wahrscheinlichkeitsberechnung standardmäßig nicht an.

Ziel dieser Arbeit soll es daher sein verschiedene Verfahren zur Wahrscheinlichkeitsberechnung (z.B. Bayesian Binning in Quantiles (BBQ)) für verschiedene statistische und maschinelle Lernverfahren zu evaluieren und die Methoden als R-Paket zu implementieren.

Literatur

- [1] Kälsch J, Bechmann LP, Heider D, Best J, Manka P, Kälsch H, Sowa JP, Moebus S, Slomiany U, Jöckel KH, Erbel R, Gerken G, Canbay A: Normal liver enzymes are correlated with severity of metabolic syndrome in a large population based cohort. *Nature Scientific Reports* 2015, 5:13058.
- [2] Dybowski JN, Riemenschneider M, Hauke S, Pyka M, Verheyen J, Hoffmann D, Heider D: Improved Bevirimat resistance prediction by combination of structural and sequence-based classifiers. *BioData Mining* 2011, 4:26.
- [3] Pyka M, Hahn T, Heider D, Krug A, Sommer J, Kircher T, Jansen A: Baseline activity predicts working memory load of preceding task condition. *Hum Brain Mapp.* 2013, 34(11):3010-22.
- [4] Sowa JP, Atmaca Ö, Kahraman A, Schlattjan M, Lindner M, Sydor S, Scherbaum N, Lackner K, Gerken G, Heider D, Arteel GE, Erim Y, Canbay A: Non-invasive separation of alcoholic and non-alcoholic liver disease with predictive modeling. *PLoS One* 2014, 9(7):e101444.

Bitte melden Sie sich bei Interesse per Email bei:
Prof. Dr. Dominik Heider
dominik.heider@uni-marburg.de

29.09.2017