

April 2024

Optimization of DNA methylation risk scores for Alzheimer´s disease

Problem Beschreibung:

Die Alzheimer-Krankheit (AD) ist eine neurodegenerative Erkrankung, die Gedächtnis, Denken und Verhalten beeinträchtigt. Um die Alzheimer-Krankheit besser zu verstehen und mögliche Medikamente und Therapien zu entwickeln, ist es wichtig, die beteiligten biologischen Pfade und Moleküle zu kennen. Die Erstellung von DNA-Methylierungsprofilen (DNAm) von Gehirnproben könnte wertvolle Informationen über die Entwicklung von Alzheimer liefern. Im Rahmen einer laufenden Zusammenarbeit mit Dr. Anke Hüls an der Emory University entwickeln wir DNAm-Risikores, die als gewichtete Summen von DNAm-Beta-Werten definiert sind, um die Neuropathologie der Alzheimer-Erkrankung vorherzusagen und Einblicke in die Stärke des Zusammenhangs zwischen dem individuellen Risikoprofil und der Alzheimer-Erkrankung zu geben.

In dieser Bachelorarbeit werden verschiedene Methoden (z.B. Random Forest; RF) zur Berechnung dieser Risikores untersucht. Sie werden optimiert, z.B. durch die Optimierung entsprechender Tuning-Parameter, um Alzheimer vorherzusagen. Der weit verbreitete statistische Lernalgorithmus RF erstellt mehrere einzelne Klassifizierungs- oder Regressionsbäume (CART), die durch einen nichtlinearen rekursiven Partitionierungsalgorithmus angepasst werden, und kombiniert sie zu einem starken Ensemble. Der Train-Test-Split ist ein Modellvalidierungsprozess, mit dem Sie simulieren können, wie Ihr Modell mit neuen Daten abschneiden würde.

Der verwendete Datensatz stammt aus der MRC London Brainbank for Neurodegenerative Disease (n=122).

References:

Hüls, A., & Czamara, D. (2020). Methodological challenges in constructing DNA methylation risk scores. *Epigenetics*, 15(1–2), 1–11. <https://doi.org/10.1080/15592294.2019.1644879>

Chen, J., Gatev, E., Everson, T., Conneely, K. N., Koen, N., Epstein, M. P., ... Hüls, A. (2023). Pruning and thresholding approach for methylation risk scores in multi-ancestry populations. *Epigenetics*, 18(1). <https://doi.org/10.1080/15592294.2023.2187172>

Lunnon K, Smith R, Hannon E, De Jager PL et al. Methyloomic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease. *Nat Neurosci* 2014 Sep; 17(9):1164-70. PMID: 25129077

Hannon E, Lunnon K, Schalkwyk L, Mill J. Interindividual methyloomic variation across blood, cortex, and cerebellum: implications for epigenetic studies of neurological and neuropsychiatric phenotypes. *Epigenetics* 2015; 10(11):1024-32. PMID: 26457534

<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE59685>

Contacts:

Prof. Dr. Katja Ickstadt (ickstast@statistik.uni-dortmund.de)

Assistant Prof. Dr. Anke Hüls (anke.huels@emory.edu)

M.Sc. Timur Tug (tug@statistik.tu-dortmund.de)

April 2024

Optimization of DNA methylation risk scores for Alzheimer's disease

Problem definition:

Alzheimer's disease (AD) is a neurodegenerative disorder that affects memory, thinking and behavior. To better understand AD and develop possible drugs and therapies, it is important to know the biological pathways and molecules involved. DNA methylation (DNAm) profiling of brain samples could provide valuable information related to the development of AD. As part of an ongoing collaboration with Dr. Anke Hüls at Emory University, we are developing DNAm risk scores defined as weighted sums of DNAm beta values to predict AD neuropathology and provide insights on the strength of association between an individual's risk profile and AD.

In this Bachelor thesis, various methods (e.g. Random Forest; RF) to calculate these risk scores will be studied. They will be optimized, e.g. by optimizing corresponding tuning parameter, in order to predict AD. The widely used statistical learning algorithm RF builds several individual classification or regression trees (CART), which are fitted by a non-linear recursive partitioning algorithm, and combines them to one strong ensemble. The train test split is a model validation process that allows you to simulate how your model would perform with new data.

The data set used comes from the MRC London Brainbank for Neurodegenerative Disease (n=122).

References:

Hüls, A., & Czamara, D. (2020). Methodological challenges in constructing DNA methylation risk scores. *Epigenetics*, 15(1–2), 1–11. <https://doi.org/10.1080/15592294.2019.1644879>

Chen, J., Gatev, E., Everson, T., Conneely, K. N., Koen, N., Epstein, M. P., ... Hüls, A. (2023). Pruning and thresholding approach for methylation risk scores in multi-ancestry populations. *Epigenetics*, 18(1). <https://doi.org/10.1080/15592294.2023.2187172>

Lunnon K, Smith R, Hannon E, De Jager PL et al. Methyloomic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease. *Nat Neurosci* 2014 Sep; 17(9):1164-70. PMID: 25129077

Hannon E, Lunnon K, Schalkwyk L, Mill J. Interindividual methyloomic variation across blood, cortex, and cerebellum: implications for epigenetic studies of neurological and neuropsychiatric phenotypes. *Epigenetics* 2015; 10(11):1024-32. PMID: 26457534

<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE59685>

Contacts:

Prof. Dr. Katja Ickstadt (ickstast@statistik.uni-dortmund.de)

Assistant Prof. Dr. Anke Hüls (anke.huels@emory.edu)

M.Sc. Timur Tug (tug@statistik.tu-dortmund.de)