Prof. Dr. Christian L. Müller
Ludwig-Maximilians-Universität München
Institut für Statistik
Ludwigstr. 33
80539 München
christian.mueller@stat.uni-muenchen.de

M.Sc. Thesis Proposal: Sparse *neural network for modality selection of multi-omic data*

Background: The integration of multiple high-dimensional data types, also referred to as multi-*omics*, offers great potential for building powerful predictive models for clinical outcomes. A key statistical challenge, however, is not just selecting predictive features, but determining the relevance of entire data modalities. Such data modalities could be an abundance table of microbes or a representation of metabolites. This problem of **modality selection** is crucial for building parsimonious, interpretable models for guiding future data collection. While methods like Group LASSO address this in a linear context, a flexible, non-linear approach remains underexplored.

Objective: The purpose of this M.Sc. thesis is to develop and utilize a **novel sparse deep learning framework** designed to perform simultaneous prediction and modality selection applied to microbiome and metabolite datasets. By embedding a statistically principled regularization scheme into a neural network architecture, we can learn to "turn off" uninformative data modalities, providing a powerful new tool for multi-omics analysis. The work sits at the intersection of deep learning and statistical regularization applied to modern biological datasets, representing a significant methodological contribution.

Plan and deliverables: A successful completion of the M.Sc. thesis requires the following computational and scientific advances. Extending and evaluating a deep neural network architecture that performs modality selection by learning to assign modality-specific importance weights for predicting a clinical outcome from microbiome and metabolome data. A write-up in thesis form and commented code on GitHub are mandatory deliverables at the end of the thesis.