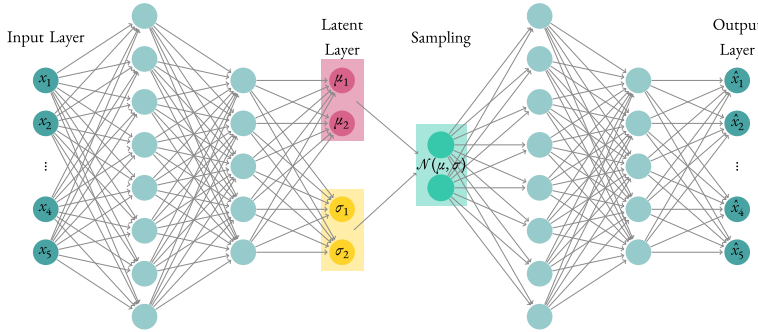
Biomolecular Dynamics
Stock Lab

www.moldyn.uni-freiburg.de
moldyn MolDynFR

Master Project

Exploring the Latent Space of Proteins with the Help of Artificial Intelligence: A Variational Approach

Molecular dynamics (MD) simulations are computational techniques used to study the behavior of, for example, proteins. Because of the large number of variables involved in a typical MD simulation of a protein, it is usually necessary to model the data in order to extract important biological information.
Typically, these models are based on the latent space, a low dimensional embedding that preserves the most relevant information. To identify such an embedding and to project the data into it, so-called information bottleneck methods are used, which pass the data through a low-dimensional bottleneck and then reconstruct the input data (see figure below).
In this project, we want to explore a recently developed approach which uses a special architecture of a neural network combined with a Bayesian approach to estimate such a latent embedding which best represents the data. [1] The project includes the implementation in Python as well as the analysis of MD simulation data of proteins in terms of a Markov state model.



Useful information

[1] Interpretable embeddings from molecular simulations using Gaussian mixture variational autoencoders; Barolgünes, Bereau, Rudinski; Machine Learning - Science and Technology, 2020

Perspective: Identification of collective variables and metastable states of protein dynamics; Sittel, Stock; Journal of Chemical Phyics, 2018

Talk to:
Prof. Gerhard Stock - stock@physik.uni-freiburg.de

universität freiburg