

Abstract

'Development of practically beneficial machine learning methods for prediction of physical-chemical and biological qualities in chemical compounds'

During the last ten years kernel based machine learning methods, like Support Vector Machines and Gaussian Processes, have established themselves as standard tools in data analysis and modeling. Unlike other methods, they are based on a solid foundation of theory.

The goal of this project is to extend kernel based machine learning methods as required by applications in the field of Chemo Informatics.

This includes the ability to quantify the confidence in each individual prediction. Model training needs to be made robust with respect to outliers in the data. For non-linear models we seek explanations of predictions & hints for optimizing the properties of individual molecules with respect to different properties.

We intend to investigate the following prototypical properties of chemical compounds: Aqueous solubility, lipophilicity, Cytochrome P450 inhibition, metabolic stability & binding to different drug targets (proteins). The emphasis of our research is both on gaining knowledge for improving machine learning methods for application in ChemInformatics and on improving these methods' foundations in learning theory.