KJM 3XXX

Machine Learning Chemistry

Motivation. At a time in which machine learning and other AI technologies are making a strong impact across multiple fields of knowledge, including Chemistry, this course will be a timely addition to the curriculum offered by the Department, attracting the attention of a broad audience of students. The course will be the natural meeting point between chemistry and the computational sciences in general, including programming. Furthermore, machine learning has a strong overlap with statistical methods such as regression, and will likely improve the student's ability to process and interpret experimental data. Outside academia, the skills developed in the course are in high demand in companies like Jotun, AstraZeneca, Kebotix, and BASF.

Brief Description. The course will provide the students in Chemistry and related disciplines with a general introduction to machine learning methods in the context of chemoinformatics. The students will learn to critically assess the quality of machine learning models, based not only on their accuracy, but also on their generality, cost, and explainability. Beyond the theoretical understanding of the fundamental concepts, the course will have a strong focus on problem-solving, which will be mostly based on programming with Python.

Learning Goals:

> Chemoinformatics representations, including molecular graphs, and similarity measures
> Use of different file types, searches in large databases, visualization of multidimensional data
> Optimization of QSAR and QSPR relationships with molecular descriptors
> Statistical methods in chemometrics, including PCA and PLS
> Prediction of properties with neural networks and Gaussian processes
> Assessment and interpretation of machine learning models
> Written and oral presentation of the results, including visual communication

Requirements. Physical chemistry and programming, both at a fundamental level.

Teaching & Evaluation. 50% theory, 50% problem-solving.

Language. English and Norwegian.

COVID Context. Easy to implement, if on-site teaching is limited by the pandemic. Both theory and problem-solving lectures can be done remotely with a computer connected to the Internet.