

# Machine Learning Challenges in Bioinformatics

Sepp Hochreiter

Johannes Kepler University Linz, Austria

<http://www.bioinf.jku.at/people/hochreiter/>



Machine learning methods like neural networks evolved in the 80's into essential bioinformatics tools that, for example, can predict the secondary and 3D structure of proteins. These days the revolution in biology and medicine concerning molecular measurement techniques creates a huge amount of high-dimensional and noisy data. For such data, machine learning provides noise reduction, feature selection, structure extraction, classification and regression in order to manage, analyze, interpret, compare, and simulate the data. In this tutorial molecular measurement data, the corresponding biomedical questions, and machine learning approaches to analyze such data will be introduced.

PART I - A prominent measurement technique is the microarray which supplies a gene expression profile in terms of mRNA concentrations – a snapshot of the cell's current state. Machine learning serves first to preprocess microarray data, then to identify relevant genes, and to predict the current or future cell states. For example, predictions in the medical context support therapy selection and diagnosis for cancer - lymphoblastic leukemia (ALL) can reliably be diagnosed by microarrays.

PART II - Sequencing methods currently rapidly increase the number of decoded genomes. However, the function and dynamics of the DNA components is not understood. To understand the DNA code, machine learning methods identify the genes and regulatory regions, detect alternative splicing, analyze gene regulation mechanism, find nucleosome positions, etc.

PART III - Further measurement techniques include x-ray, NMR, mass spectrometry, atomic force microscopy, fluorescence microscopy, protein arrays, peptide arrays, etc. For example, peptide arrays are able to detect antibodies which are specific for certain diseases like cancer. To identify peptides which bind specific antibodies, feature selection is applied to random peptide arrays. Thereafter a classifier is constructed which is based on the selected peptides and which can detect the disease.

.....  
Presently owns the position of Head of Institute of Bioinformatics, Johannes Kepler University Linz. Research Topics: Microarrays, Protein structure, SNP-Data, Alternative splicing, Medical approach of bioinformatics, Nucleosome positioning, Peptide arrays, Viral genomes, Sequence analysis, Machine Learning, Statistics. Graduated in 1992 at the Technische Universität München, in computer science and mathematics and at FernUniversität Hagen in Mathematics. In 1994-1999 at the Technische Universität München obtained a PhD in computer science. Employments in 1994-1999: Research Associate, Munich University of Technology, and, in 1992-1994: Head of a group at Allianz AG München: Project "*Schadenußendienstdatei*".