

Sparse Proteomics Analysis

Feature Selection for Mass Spectrometry Data via 1-Bit Compressed Sensing

Martin Genzel

TECHNISCHE UNIVERSITÄT BERLIN
INSTITUT FÜR MATHEMATIK

email: genzel@math.tu-berlin.de

Abstract

Tumor diseases as cancer rank among the most frequent causes of death in Western countries. To date, most of the underlying pathological mechanisms are not completely understood, and therefore, early diagnostics is the best available option to improve patient survival. It is well-known that many of the diseases-causing biological processes are controlled by proteins. The set of all proteins in the human body, a so-called *proteome*, can be detected in the blood stream, for instance, by *mass spectrometry*. It turned out that the diseases-indicating features in the resulting high-dimensional data are usually sparse. Thus, it is a major goal to find a small set of *fingerprints* enabling us to classify healthy and diseased people.

In this presentation, we will see that the recently introduced framework of *1-Bit Compressed Sensing* is naturally applicable to this kind of problem. Combined with a *Gaussian dictionary*, this approach indeed leads to a robust and tractable algorithm providing sparse and interpretable feature sets. The related reduction of dimension finally allows us to apply various methods from *machine learning* in order to achieve an appropriate classification of the patients.

Keywords: High-Dimensional Data Analysis, Proteomics Analysis, 1-Bit Compressed Sensing, Gaussian Dictionaries, Machine Learning