

Targeted NMF for Tumor Classification

Y. C. Hernandez¹, T. Boskamp², and P. Maass³

Abstract: Matrix Assisted Laser Desorption Imaging (MALDI) is an analytical method used for the investigation of biological tissues. One of its applications is to classify different tumor types to support histopathological diagnosis. For the obtention of robust, not overfitted classifiers, an appropriate feature selection is an essential step. In the context of MALDI imaging Non-negative Matrix Factorization (NMF) has been successfully used to obtain a low-rank spectral decomposition of the data. Under mild assumptions such factorizations capture essential characteristics of the underlying proteins generating the data. Moreover some type of tumors can be associated with the presence of certain proteins. Obtaining a precise distribution of those biomarkers in the analyzed tissue from the MALDI data can be therefore relevant to the classification task. Unfortunately a standard NMF can not provide a reliable approximation of the ionization and distribution of a specific protein. In this work we propose an extension of NMF where information related to the peptide composition of targeted proteins and their most probable peptide ionization is included in the optimization model related to the decomposition. First promising results obtained with this novel approach using synthetic and real world data are shown.

^{1,2,3} Center for Industrial Mathematics
University of Bremen
Bibliothekstrae 1, 28359 Bremen, Germany
{*yovany, tboskamp, pmaass*}@math.uni-bremen.de