Non-negative Matrix Factorization for Binary Data

We propose the Logistic Non-negative Matrix Factorization for decomposition of binary data. Binary data are frequently generated in e.g. text analysis, sensory data, market basket data etc. A common method for analysing non-negative data is the Non-negative Matrix Factorization, though this is in theory not appropriate for binary data, and thus we propose a novel Non-negative Matrix Factorization based on the logistic link function. Furthermore we generalize the method to handle missing data. The formulation of the method is compared to a previously proposed method (Tome et al., 2015). We compare the performance of the Logistic Non-negative Matrix Factorization to Least Squares Non-negative Matrix Factorization and Kullback-Leibler (KL) Non-negative Matrix Factorization on sets of binary data: a synthetic dataset, a set of student comments on their professors collected in a binary term-document matrix and a sensory dataset. We find that choosing the number of components is an essential part in the modelling and interpretation, that is still unresolved.

General information
Publication status: Published
Organisations: Department of Applied Mathematics and Computer Science, Statistics and Data Analysis
Contributors: Larsen, J. S., Clemmensen, L. K. H.
Number of pages: 9
Publication date: 2015
Peer-reviewed: Yes
Event: Paper presented at 7th International Conference on Knowledge Discovery and Information Retrieval (KDIR 2015), Lisbon, Portugal.
Keywords: Non-negative Matrix Factorization, Binary Data, Binary Matrix Factorization, Text Modelling
Source: PublicationPreSubmission
Source-ID: 118545190
Research output: Contribution to conference › Paper – Annual report year: 2015 › Research › peer-review