

Zero-inflated negative-binomial NMF

Hiroyasu Abe

Kyoto University

E-mail: hiroabe@kuhp.kyoto-u.ac.jp

Abstract: co-authors are Hiroyasu Abe (Kyoto University, Kyoto, Japan) and Hiroshi Yadohisa (Doshisha, University, Kyoto, Japan) Nonnegative matrix factorization (NMF) is a matrix decomposition technique to capture hidden structures in a nonnegative data matrix, the entries of which are all nonnegative, for example, multivariate count data. The solution of NMF differs depending on what probability distribution is assumed. A Poisson distribution is the most commonly used probability distribution for modeling count data in NMF. Moreover, a new NMF method has been proposed recently based on the negative binomial distribution, which is compatible with overdispersed count data. However, these NMF methods do not account for the zero-inflated case, wherein the data matrix has many zeros. The zero-inflated Poisson distribution is a solution for the zero-inflated case, and NMF based on such a distribution has already been proposed. However, the NMF method to take into account zero-inflated case using the negative binomial distribution is not yet proposed. In this study, we propose the new NMF method using zero-inflated negative binomial distribution (ZINBNMF) to consider the overdispersion and zero-inflation of count data. We evaluate its performance by numerical simulation.