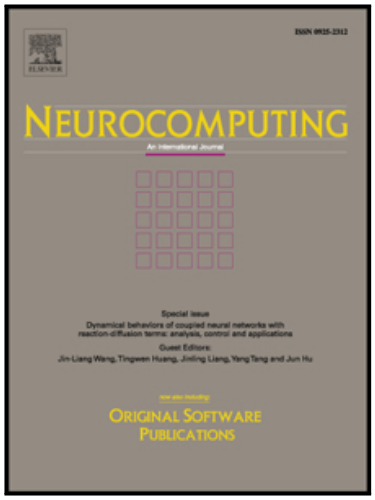
Communicated by Dr Xin Luo

Accepted Manuscript



Upper Bound of Bayesian Generalization Error in Non-Negative Matrix Factorization

Naoki HAYASHI, Sumio WATANABE

PII: S0925-2312(17)30845-7
DOI: 10.1016/j.neucom.2017.04.068
Reference: NEUCOM 18442

To appear in: *Neurocomputing*

Received date: 15 December 2016
Revised date: 22 February 2017
Accepted date: 27 April 2017

Please cite this article as: Naoki HAYASHI, Sumio WATANABE, Upper Bound of Bayesian Generalization Error in Non-Negative Matrix Factorization, *Neurocomputing* (2017), doi: 10.1016/j.neucom.2017.04.068

Available online at www.sciencedirect.com
ScienceDirect
Neurocomputing 00 (2017) 1–18
Neurocomputing
www.elsevier.com/locate/procedia

Upper Bound of Bayesian Generalization Error in Non-Negative Matrix Factorization

Naoki HAYASHI[a], Sumio WATANABE[b]

[a] *Department of Mathematical and Computing Science*
Tokyo Institute of Technology
Mailbox G5-19, 4259 Nagatsuta, Midori-ku, Yokohama, Kanagawa, 226–8502, Japan
hayashi.n.ag@m.titech.ac.jp
[b] *Department of Mathematical and Computing Science*
Tokyo Institute of Technology
Mailbox G5-19, 4259 Nagatsuta, Midori-ku, Yokohama, Kanagawa, 226–8502, Japan
swatanab@c.titech.ac.jp

Abstract

Non-negative matrix factorization (NMF) is a new knowledge discovery method that is used for text mining, signal processing, bioinformatics, and consumer analysis. However, its basic property as a learning machine is not yet clarified, as it is not a regular statistical model, resulting that theoretical optimization method of NMF has not yet established. In this paper, we study the real log canonical threshold of NMF and give an upper bound of the generalization error in Bayesian learning. The results show that the generalization error of the matrix factorization can be made smaller than regular statistical models if Bayesian learning is applied.

Keywords: non-negative matrix factorization (NMF), real log canonical threshold (RLCT), Bayesian learning

1. Introduction

Recently, non-negative matrix factorization (NMF) [1, 2] has been applied to text mining [3], signal processing [4, 5, 6], bioinformatics [7], and consumer analysis [8]. Experiments has shown that a new knowledge discovery method is derived by NMF, however, its mathematical property as a learning machine is not yet clarified, since it is not a regular statistical model. A statistical model is called regular if a function from a parameter to a probability density function is one-to-one and if the likelihood function can be approximated by a Gaussian function.

It is proved that, if a statistical model is regular and if a true distribution is realizable by a statistical model, then the generalization error is asymptotically equal to $d/(2n)$, where d, n, and the generalization error are the dimension of the parameter, the sample size, and the expected Kullback-Leibler divergence of the true distribution and the estimated learning machine, respectively. However, the statistical model used in NMF is not regular because the map from a parameter to a probability density function is not injective. As a result, its generalization error is