Accepted Manuscript

Continuum directions for supervised dimension reduction

Sungkyu Jung

PII: S0167-9473(18)30075-6
DOI: https://doi.org/10.1016/j.csda.2018.03.015
Reference: COMSTA 6587

To appear in: *Computational Statistics and Data Analysis*

Received date : 24 March 2017
Revised date : 5 March 2018
Accepted date : 21 March 2018

Please cite this article as: Jung S., Continuum directions for supervised dimension reduction. *Computational Statistics and Data Analysis* (2018), https://doi.org/10.1016/j.csda.2018.03.015

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

*Manuscript
Click here to view linked References

Continuum directions for supervised dimension reduction

Sungkyu Jung

Department of Statistics, University of Pittsburgh, Pittsburgh, PA 15260, U.S.A.

Abstract

Dimension reduction of multivariate data supervised by auxiliary information is considered. A series of basis for dimension reduction is obtained as minimizers of a novel criterion. The proposed method is akin to continuum regression, and the resulting basis is called continuum directions. With a presence of binary supervision data, these directions continuously bridge the principal component, mean difference and linear discriminant directions, thus ranging from unsupervised to fully supervised dimension reduction. High-dimensional asymptotic studies of continuum directions for binary supervision reveal several interesting facts. The conditions under which the sample continuum directions are inconsistent, but their classification performance is good, are specified. While the proposed method can be directly used for binary and multi-category classification, its generalizations to incorporate any form of auxiliary data are also presented. The proposed method enjoys fast computation, and the performance is better or on par with more computer-intensive alternatives.

Keywords: continuum regression, dimension reduction, linear discriminant analysis, high-dimension, low-sample-size (HDLSS), maximum data piling, principal component analysis
2000 MSC: 60K35

1. Introduction

In modern complex data, it becomes increasingly common that multiple data sets are available. We consider the data situation where a supervised

Email address: sungkyu@pitt.edu (Sungkyu Jung)

Download English Version:

https://daneshyari.com/en/article/6868664

Download Persian Version:

https://daneshyari.com/article/6868664

Daneshyari.com