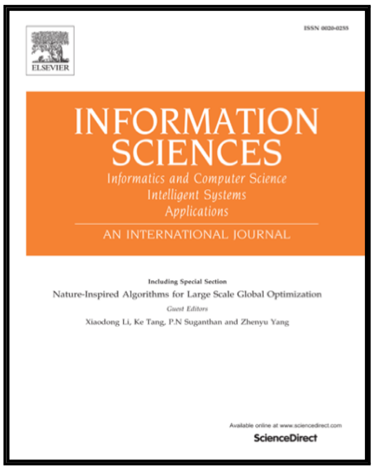
Accepted Manuscript



Optimizing Cluster Structures with Inner Product Induced Norm based Dissimilarity Measures: Theoretical Development and Convergence Analysis

Arkajyoti Saha, Swagatam Das

PII: S0020-0255(16)30628-4
DOI: 10.1016/j.ins.2016.08.058
Reference: INS 12461

To appear in: *Information Sciences*

Received date: 3 December 2015
Revised date: 28 June 2016
Accepted date: 15 August 2016

Please cite this article as: Arkajyoti Saha, Swagatam Das, Optimizing Cluster Structures with Inner Product Induced Norm based Dissimilarity Measures: Theoretical Development and Convergence Analysis, *Information Sciences* (2016), doi: 10.1016/j.ins.2016.08.058

Optimizing Cluster Structures with Inner Product Induced Norm based Dissimilarity Measures: Theoretical Development and Convergence Analysis

Arkajyoti Saha[a], Swagatam Das[b,*]

[a] Stat-Math Unit, Indian Statistical Institute, 203 B.T.Road, Kolkata-700108, W.B., INDIA.
[b] Electronics and Communication Sciences Unit, Indian Statistical Institute, 203 B.T.Road, Kolkata-700108, W.B., INDIA.

Abstract

Dissimilarity measures play a key role in exploring the inherent cluster structure of the data for any partitional clustering algorithm.Commonly used dissimilarity functions for clustering purpose are so far confined to the Euclidean, exponential and Mahalanobish distances. In this article we develop generalized algorithms to solve the partitional clustering problems formulated with a general class of Inner Product Induced Norm (IPIN) based dissimilarity measures. We provide an in-depth mathematical analysis of the underlying optimization framework and analytically address the issue of existence of a solution and its uniqueness. In absence of a closed form solution, we develop a fast stochastic gradient descent algorithm and the Minimization by Incremental Surrogate Optimization (MISO) algorithm (in case of constrained optimization) with exponential convergence rate to obtain the solution. We carry out a convergence analysis of the fuzzy and *k*-means clustering algorithms with the IPIN based dissimilarity measures and also establish how these algorithms guarantee convergence to a stationary point. In addition, we investigate the nature of the stationary point. Novelty of the paper lies in the introduction of a generalized class of divergence measures, development of fuzzy and *k*-means clustering algorithms with the general class of divergence measures and a thorough convergence analysis of the developed algorithms.

Keywords: Dissimilarity measures, fuzzy covariance matrix, inner product induced norm, *k*-means and fuzzy clustering, stochastic gradient descent, MISO, convergence analysis.

1. Introduction

Clustering is the unsupervised partitioning of a collection of patterns (usually represented as a vector of measurements or a point in some multi-dimensional space) based on certain dissimilarity measures in such a way, that the patterns belonging to the same cluster (same label) are more similar to each other, than they are to the patterns belonging to the other clusters [38]. The absence of any kind of training dataset i.e. unsupervised nature of the problem makes this method

*Corresponding author
Email addresses: arkajyotisaha93@gmail.com (Arkajyoti Saha), swagatam.das@isical.ac.in (Swagatam Das)