

Project Proposal: Convex Latent Variable

Anonymous Author(s)

Affiliation

Address

email

Abstract

Add abstract here..

1 Problem Formulation

1.1 Motivation

Given a set of data entities $\mathbf{x}_1, \dots, \mathbf{x}_n, \dots, \mathbf{x}_N$, we wish to automatically cluster these entities without specifying the number of cluster centroids. One possible approach is to regard these input entities as potential cluster candidates and evaluate the "belonging matrix" $\mathbf{w}_{n \times n}$, consisting of latent variables w_{nk} indicating the extent of one entity \mathbf{x}_n explained by one potential cluster centroid candidate μ_k .

1.2 Overall Optimization Goal

Formally, our goal is to achieve clustering with group-lasso regularization. Hence, we formulate the target problem in terms of notations introduced above as follows:

$$\underset{x}{\text{minimize}} \quad J(x) = \frac{1}{2} \sum_n \sum_k w_{nk} \|\mathbf{x}_n - \mu_k\|^2 + \lambda \sum_k \max_n |w_{nk}| \quad (1)$$

$$\text{subject to} \quad \forall n, \sum_k w_{nk} \leq 1 \quad (2)$$

$$\forall n, k, w_{nk} \geq 0 \quad (3)$$

Every entity must be assigned to some centroid with prob 1. Every entity must have non-negative belonging to each centroid candidate.

Note that the group-lasso can either be one or zero in the context that we pick up the most promising centroid candidates from provided entities.

1.3 Processing Techniques

Acknowledgments

References