
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 . In this project, we propose a method to compute the sparse latent variable with faster convergence by decomposing the objective function derived by ADMM into two separate optimization task.

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:

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

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

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

The limiting conditions listed above are jointly called simplex constraint. Every entity must be assigned to some centroid with prob 1. Every entity must have non-negative belonging index 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 ADMM

We can rewrite the initial optimization goal as the following primal problem:

$$\begin{aligned}
& \underset{\mathbf{w}_1, \mathbf{w}_2, \mathbf{z}}{\text{Minimize}} && \frac{1}{2} \sum_n \sum_k w_{nk} \|\mathbf{x}_n - \boldsymbol{\mu}_k\|^2 + \frac{\rho}{2} \|\mathbf{w}_1 - \mathbf{z}\|^2 \\
& && + \lambda \sum_k \max_n |w_{nk}| + \frac{\rho}{2} \|\mathbf{w}_2 - \mathbf{z}\|^2 \\
& \text{subject to} && \mathbf{w}_1 = \mathbf{z}, \mathbf{w}_2 = \mathbf{z}
\end{aligned} \tag{4}$$

Note that we separate \mathbf{w} into two variables, \mathbf{w}_1 and \mathbf{w}_2 respectively. This is to prepare it for latter optimization decomposition.

By dual decomposition, the problem can be further transformed as follows:

$$\begin{aligned}
& \underset{\mathbf{w}_1, \mathbf{w}_2, \mathbf{z}}{\text{Minimize}} && \underset{\mathbf{y}_1, \mathbf{y}_2}{\text{Maximize}} && \frac{1}{2} \sum_n \sum_k w_{nk} \|\mathbf{x}_n - \boldsymbol{\mu}_k\|^2 + \mathbf{y}_1^T (\mathbf{w}_1 - \mathbf{z}) + \frac{\rho}{2} \|\mathbf{w}_1 - \mathbf{z}\|^2 \\
& && && + \lambda \sum_k \max_n |w_{nk}| + \mathbf{y}_2^T (\mathbf{w}_2 - \mathbf{z}) + \frac{\rho}{2} \|\mathbf{w}_2 - \mathbf{z}\|^2
\end{aligned} \tag{6}$$

We continue by incorporating duality and derive the following optimization task:

$$\begin{aligned}
& \underset{\mathbf{y}_1, \mathbf{y}_2}{\text{Maximize}} && \underset{\mathbf{w}_1, \mathbf{w}_2, \mathbf{z}}{\text{Minimize}} && \frac{1}{2} \sum_n \sum_k w_{nk} \|\mathbf{x}_n - \boldsymbol{\mu}_k\|^2 + \mathbf{y}_1^T (\mathbf{w}_1 - \mathbf{z}) + \frac{\rho}{2} \|\mathbf{w}_1 - \mathbf{z}\|^2 \\
& && && + \lambda \sum_k \max_n |w_{nk}| + \mathbf{y}_2^T (\mathbf{w}_2 - \mathbf{z}) + \frac{\rho}{2} \|\mathbf{w}_2 - \mathbf{z}\|^2
\end{aligned} \tag{7}$$

1.4 Optimization Decomposition

The first optimization task:

$$\underset{\mathbf{w}_1}{\text{Minimize}} \quad \frac{1}{2} \sum_n \sum_k w_{nk} \|\mathbf{x}_n - \boldsymbol{\mu}_k\|^2 + \mathbf{y}_1^T \mathbf{w}_1 + \frac{\rho}{2} \|\mathbf{w}_1 - \mathbf{z}\|^2 \tag{8}$$

$$\text{subject to} \quad \forall n, \sum_k w_{nk} \leq 1 \tag{9}$$

$$\forall n, k, w_{nk} \geq 0 \tag{10}$$

This subproblem can be solved by Frank-Wolf Algorithm.

The second optimization task:

$$\underset{\mathbf{w}_2}{\text{Minimize}} \quad \lambda \sum_k \max_n |w_{nk}| + \mathbf{y}_2^T \mathbf{w}_2 + \frac{\rho}{2} \|\mathbf{w}_2 - \mathbf{z}\|^2 \tag{11}$$

$$\tag{12}$$

The closed-form solution of second subproblem can be found through Blockwise Coordinate Descent Procedures (Han Liu.).

1.5 Update Rules

Acknowledgments

References