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 $\mathbf{x}_1, ..., \mathbf{x}_n, ..., \mathbf{x}_N$, 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 $\boldsymbol{\mu}_k$. 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:

$$\underset{\mathbf{w}}{\text{Minimize}} \quad \frac{1}{2}\sum_n \sum_k w_{nk}||\mathbf{x}_n - \boldsymbol{\mu}_k||^2 + \lambda \sum_k \max_n |w_{nk}| \tag{1}$$

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

$$\forall n,\ k,\ w_{nk} \geq 0 \tag{3}$$

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:

$$
\begin{aligned}
\underset{\mathbf{w}_1,\mathbf{w}_2,\mathbf{z}}{\text{Minimize}} \quad & \frac{1}{2}\sum_n \sum_k w_{nk}||\mathbf{x}_n - \boldsymbol{\mu}_k||^2 + \frac{\rho}{2}||\mathbf{w}_1 - \mathbf{z}||^2 \\
& \quad + \lambda \sum_k \max_n |w_{nk}| + \frac{\rho}{2}||\mathbf{w}_2 - \mathbf{z}||^2 && (4) \\
\text{subject to} \quad & \mathbf{w}_1 = \mathbf{z},\ \mathbf{w}_2 = \mathbf{z} && (5)
\end{aligned}
$$

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:

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

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

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

1.4 Optimization Decomposition

The first optimization task:

$$
\begin{aligned}
\underset{\mathbf{w}_1}{\text{Minimize}} \quad & \frac{1}{2}\sum_n \sum_k w_{nk}||\mathbf{x}_n - \boldsymbol{\mu}_k||^2 + \mathbf{y}_1^T\mathbf{w}_1 + \frac{\rho}{2}||\mathbf{w}_1 - \mathbf{z}||^2 && (8) \\
\text{subject to} \quad & \forall n,\ \sum_k w_{nk} \leq 1 && (9) \\
& \forall n,\ k,\ w_{nk} \geq 0 && (10)
\end{aligned}
$$

This subproblem can be solved by Frank-Wolf Algorithm.

The second optimization task:

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

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

1.5 Overall Solution Procedure

Repeat

1. Resolve $\mathbf{w}_1$ and $\mathbf{w}_2$ in separate subproblem formulated in section 1.4.

2. Update the combination indicator $\mathbf{z}$ by $\mathbf{z} = \frac{1}{2}(\mathbf{w}_1 + \mathbf{w}_2)$

3. Refine the multiplier $\mathbf{y}_1$ and $\mathbf{y}_2$ with gradient update rule:

$$
\begin{aligned}
\mathbf{y}_1 &= \mathbf{y}_1 - \alpha \cdot (\mathbf{w}_1 - \mathbf{z}) && (12) \\
\mathbf{y}_2 &= \mathbf{y}_2 - \alpha \cdot (\mathbf{w}_2 - \mathbf{z}) && (13)
\end{aligned}
$$

Until Convergence.

1.6 Resolve first subproblem: Frank-Wolf Algorithm

1.6.1 Single-variable case

In the context of only one input instance, the optimization problem can be written as

$$\underset{w}{\text{Minimize}} \quad \frac{1}{2}w \cdot d^2 + y \cdot w + \frac{\rho}{2}||w - z||^2 \tag{14}$$

$$\text{subject to} \quad 0 \leq w \leq 1 \tag{15}$$

Note that the distance d above should be zero. We can further rewrite above optimization task as

$$\underset{w}{\text{Minimize}} \quad \frac{\rho}{2}w^2 + (-\rho z + y + \frac{1}{2}d^2)w + z^2 \tag{16}$$

$$\text{subject to} \quad 0 \leq w \leq 1 \tag{17}$$

As we can see the critical point of objective function is

$$-\frac{-\rho z + y + \frac{1}{2}d^2}{\rho} = z - \frac{1}{\rho}y - \frac{d^2}{2\rho} \tag{18}$$

1.7 Resolve second subproblem: Blockwise Coordinate Descent Procedures

According to Blockwise Coordinate Descent Procedures (Han Liu.), the way to resolve the (11) optimization problem is first to seek solution $\overline{\mathbf{w}}$ without group-lasso (19) as shown below.

$$\underset{\mathbf{w}_2}{\text{Minimize}} \quad \mathbf{y}_2^T\mathbf{w}_2 + \frac{\rho}{2}||\mathbf{w}_2 - \mathbf{z}||^2 \tag{19}$$

Then ultimately find the closed-form solution in terms of $\overline{\mathbf{w}}$. In practice, we resolve it column by column. The formula for closed-form solution $\mathbf{w}^*$ is given as follows:

$$\mathbf{w}^* = ... \tag{20}$$

where $m^* = \underset{m}{\text{argmax}} \frac{1}{m}\big(\sum_{i'}^{m} |\mathbf{w}^{*(k_{i'})}| - \lambda\big)$.

2 Application on Topic Model

2.1

2.2

Acknowledgments

References