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:

$$\underset{\mathbf{w}_1,\mathbf{w}_2,\mathbf{z}}{\text{Minimize}} \quad \frac{1}{2}\sum_n\sum_k w_{nk}||\mathbf{x}_n - \boldsymbol{\mu}_k||^2 + \frac{\rho}{2}||\mathbf{w}_1 - \mathbf{z}||^2$$

$$+ \lambda\sum_k \max_n |w_{nk}| + \frac{\rho}{2}||\mathbf{w}_2 - \mathbf{z}||^2 \tag{4}$$

$$\text{subject to} \quad \mathbf{w}_1 = \mathbf{z}, \ \mathbf{w}_2 = \mathbf{z} \tag{5}$$

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:

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

$$+ \lambda\sum_k \max_n |w_{nk}| + \mathbf{y}_2^T(\mathbf{w}_2 - \mathbf{z}) + \frac{\rho}{2}||\mathbf{w}_2 - \mathbf{z}||^2 \tag{6}$$

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

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

$$+ \lambda\sum_k \max_n |w_{nk}| + \mathbf{y}_2^T(\mathbf{w}_2 - \mathbf{z}) + \frac{\rho}{2}||\mathbf{w}_2 - \mathbf{z}||^2 \tag{7}$$

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:

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

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:

$$\mathbf{y}_1 = \mathbf{y}_1 - \alpha \cdot (\mathbf{w}_1 - \mathbf{z}) \tag{12}$$

$$\mathbf{y}_2 = \mathbf{y}_2 - \alpha \cdot (\mathbf{w}_2 - \mathbf{z}) \tag{13}$$

Until Convergence.

2

108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161

1.6 Resolve first subproblem: Frank-Wolf Algorithm

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 (14) as shown below.

$$\underset{\mathbf{w}_2}{\text{Minimize}} \quad \mathbf{y}_2^T \mathbf{w}_2 + \frac{\rho}{2}||\mathbf{w}_2 - \mathbf{z}||^2 \tag{14}$$

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:

$$\mathbf{w}^* = ... \tag{15}$$

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

2 Application on Topic Model

2.1 Problem Formulation

In this topic model, there are four main semantics supposed to be captured.

First key point is that all words in any document should be related to one topic. This can be regarded as resources (words) given out to all receivers (topics). And intuitively, this term support the value of components within w. Formally, this is represented by equality constraints.

Secondly, too many hidden topics are supposed to be penalized. Intuitively, this objective aimed at reducing total number of topics. Formally, one regularization term is used as follows:

$$\lambda_g \sum_k \max_n |w_{nk}| \tag{16}$$

Thirdly, too many topics involved by each document should be penalized. However, the situation of no topics to explain one document is non-existent since we enforce one constraint that all words must be related to certain topics with total explanation. Hence, this objective intuitively aimed at reducing topic distribution within one document. To capture this semanticcs, we use another regularizer:

$$\lambda_l \sum_d \sum_k \max_{n \in d} |w_{nk}| \tag{17}$$

Fourthly, too many vocabularies involved by each topic ought to be penalized. This would increase the number of hidden topics within this model. It can be intuitively seen as an support of topics number or cause of matrix extension of w. One extra regularizer is employed:

$$\lambda_b \sum_k \sum_{w \in voc(k)} \max_{word(n)=w} |w_{nk}| \tag{18}$$

To acquire small penalty from the last regularization, all documents that are related to certain topic should be attributed to a narrow set of vocabularies.

2.2 Objectives

$$\underset{w}{\text{Minimize}} \quad \lambda_g \sum_k \max_n |w_{nk}| + \lambda_l \sum_d \sum_k \max_{n \in d} |w_{nk}| + \lambda_b \sum_k \sum_{w \in voc(k)} \max_{word(n)=w} |w_{nk}| \tag{19}$$

$$\text{subject to} \quad \forall n, \, k, \, 0 \le w_{nk} \le 1 \tag{20}$$

$$\forall n, \, \sum_k w_{nk} = 1 \tag{21}$$

2.3 Optimization Decomposition

The first optimization task:

$$\underset{\mathbf{w}_1}{\text{Minimize}} \quad r\left|1 - \sum_k w_{nk}\right| + \mathbf{y}_1^T \mathbf{w}_1 + \frac{\rho}{2}||\mathbf{w}_1 - \mathbf{z}||^2 \tag{22}$$

$$\text{subject to} \quad \forall n, \, \sum_k w_{nk} \le 1 \tag{23}$$

$$\forall n, \, k, \, w_{nk} \ge 0 \tag{24}$$

where r is a arbitrary value to enforce the resource to be fully distributed. Note that this subproblem can be solved by Frank-Wolf Algorithm.

The second optimization task:

$$\underset{\mathbf{w}_2}{\text{Minimize}} \quad \lambda_g \sum_k \max_n |w_{nk}| + \mathbf{y}_2^T \mathbf{w}_2 + \frac{\rho}{2}||\mathbf{w}_2 - \mathbf{z}||^2 \tag{25}$$

4

The third optimization task:

$$\underset{\mathbf{w}_3}{\text{Minimize}} \quad \lambda_l \sum_d \sum_k \max_{n \in d} |w_{nk}| + \mathbf{y}_3^T \mathbf{w}_3 + \frac{\rho}{2} ||\mathbf{w}_3 - \mathbf{z}||^2 \tag{26}$$

The fourth optimization task:

$$\underset{\mathbf{w}_4}{\text{Minimize}} \quad \lambda_b \sum_k \sum_{w \in voc(k)} \max_{word(n)=w} |w_{nk}| + \mathbf{y}_4^T \mathbf{w}_4 + \frac{\rho}{2} ||\mathbf{w}_4 - \mathbf{z}||^2 \tag{27}$$

The closed-form solution of 2-4 subproblems can be found through Blockwise Coordinate Descent Procedures (Han Liu.).

2.4 Overall Solution Procedure

Repeat

1. Resolve $\mathbf{w}_1$, $\mathbf{w}_2$, $\mathbf{w}_3$, $\mathbf{w}_4$ in separate subproblem formulated in section 2.3.

2. Update the combination indicator $\mathbf{z}$ by $\mathbf{z} = \frac{1}{4}(\mathbf{w}_1 + \mathbf{w}_2 + \mathbf{w}_3 + \mathbf{w}_4)$

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

$$\mathbf{y}_1 = \mathbf{y}_1 - \alpha \cdot (\mathbf{w}_1 - \mathbf{z}) \tag{28}$$
$$\mathbf{y}_2 = \mathbf{y}_2 - \alpha \cdot (\mathbf{w}_2 - \mathbf{z}) \tag{29}$$
$$\mathbf{y}_3 = \mathbf{y}_3 - \alpha \cdot (\mathbf{w}_3 - \mathbf{z}) \tag{30}$$
$$\mathbf{y}_4 = \mathbf{y}_4 - \alpha \cdot (\mathbf{w}_4 - \mathbf{z}) \tag{31}$$

Until Convergence.

2.5 Implementation Details

To handle large scale data and manage extensibility, we implement a new data structure - extensible sparse matrice.

5

Acknowledgments

References