HOW TO SOLVE THE PROBLEMS IN OUR PROJECT?

Tikhonov regularization: consider the item-item graph Tikhonov regularization

$$\min_{\mathbf{X}\in\mathbb{R}^{U\times I}} \|\mathbf{Y}-\mathbf{X}\|_F^2+\mu\, tr\left(\mathbf{X}^\top\mathbf{L}\mathbf{X}\right) \tag{30}$$

where $\|\cdot\|_F$ is the Frobenius norm and matrix $\mathbf{X}\in\mathbb{R}^{U\times I}$ is the full rating matrix.

Equivalently, for the ith item, $i=1,\ldots,I$, the regularization problem has the form to solve the ratings $\mathbf{x}_i\in\mathbb{R}^U$ for all users

$$\min_{\mathbf{x}_i\in\mathbb{R}^U} \|\mathbf{y}_i-\mathbf{x}_i\|_2^2+\mu\,\mathbf{x}_i^\top\mathbf{L}\mathbf{x}_i. \tag{31}$$

Let us write the objective out as the following form (subscript i omitted)

$$\begin{aligned}&(\mathbf{y}-\mathbf{x})^\top(\mathbf{y}-\mathbf{x})+\mu\,\mathbf{x}^\top\mathbf{L}\mathbf{x}\\ =&(\mathbf{x}^\top\mathbf{x}-2\mathbf{y}^\top\mathbf{x}+\mathbf{y}^\top\mathbf{y})+\mu\,\mathbf{x}^\top\mathbf{L}\mathbf{x}\end{aligned} \tag{32}$$

which is a convex function w.r.t., $\mathbf{x}$. We can find its gradient as follows

$$\nabla_{\mathbf{x}} = 2\mathbf{x}-2\mathbf{y}+2\mu\mathbf{L}\mathbf{x}. \tag{33}$$

Since above gradient has a zeros, i.e., $\nabla_{\mathbf{x}}=\mathbf{0}\in\mathbb{R}^U$ has a closed-form solution, we can find the optimum $\mathbf{x}^*$ as

$$\begin{aligned}\nabla_{\mathbf{x}}&=\mathbf{0}\\ (\mathbf{I}+\mu\mathbf{L})\mathbf{x}&=\mathbf{y}\\ \mathbf{x}^*&=(\mathbf{I}+\mu\mathbf{L})^{-1}\mathbf{y}.\end{aligned} \tag{34}$$

matrix form solution

Intuitively, the solution in the matrix form should be

$$\mathbf{X}^*=(\mathbf{I}+\mu\mathbf{L})^{-1}\mathbf{Y}. \tag{35}$$

But can you find this solution starting from the objective function of the matrix form? Try it with matrix cookbook.

other regularizers

Depends on if we have a closed-form gradient? closed-form solution from setting the gradient as zero? then we choose the way of solving the problem

- adjacency based 2-norm variation measure $\|\mathbf{x}-\mathbf{A}_n\mathbf{x}\|_2^2$:
- total variation 1-norm $\|\mathbf{x}-\mathbf{A}_n\mathbf{x}\|_1$: check the ISTA algorithm based on soft-thresholding
- SLIM method has two regularizers: $\|\mathbf{w}\|_2^2+\|\mathbf{w}\|_1$

In the case of other regularizers, you could replace the term $tr\left(\mathbf{X}^\top\mathbf{L}\mathbf{X}\right)$ in (30) or $\mathbf{x}_i^\top\mathbf{L}\mathbf{x}_i$ in (31) by the corresponding regularizers. For instance, the total variation 1-norm one would have the form in the vector fashion:

$$\min_{\mathbf{x}_i\in\mathbb{R}^U} \|\mathbf{y}_i-\mathbf{x}_i\|_2^2+\mu\,\|\mathbf{x}_i-\mathbf{A}_n\mathbf{x}_i\|_1. \tag{36}$$

In the matrix form, there is an equivalent form too as follows:

$$\min_{\mathbf{X}\in\mathbb{R}^{U\times I}} \|\mathbf{Y}-\mathbf{X}\|_2^2+\mu\,\|\mathbf{X}-\mathbf{A}_n\mathbf{X}\|_1. \tag{37}$$

Here $\|\mathbf{X}-\mathbf{A}_n\mathbf{X}\|_1$ is hard to handle, which is also one of the reasons why we recommend to solve the problem in the vector form.

For another example, the regularizer $\|\mathbf{x}-\mathbf{A}_n\mathbf{x}\|_2^2$. The problem of the vector form is

$$\min_{\mathbf{x}_i\in\mathbb{R}^U} \|\mathbf{y}_i-\mathbf{x}_i\|_2^2+\mu\,\|\mathbf{x}_i-\mathbf{A}_n\mathbf{x}_i\|_2^2. \tag{38}$$

In the matrix form, it is

$$\min_{\mathbf{X}\in\mathbb{R}^{U\times I}} \|\mathbf{Y}-\mathbf{X}\|_2^2+\mu\,\|\mathbf{X}-\mathbf{A}_n\mathbf{X}\|_F. \tag{39}$$

Basically, $\|\mathbf{X}\|_F$ is a matrix version of the vector norm $\|\mathbf{x}\|_2$.

ℓ_1 norm

Let us have a closer look at the ℓ_1 norm, the uth element

$$(\|\mathbf{x}\|_1)_u=\begin{cases}x_u, & \text{if } x_u>0\\ -x_u, & \text{if } x_u<0\\ 0, & \text{if } x_u=0.\end{cases} \tag{40}$$

What about its gradient? Technically, we cannot find the gradient, why? Check the plot of $|x|$, at $x=0$ there are indefinite gradients. Thus, so-called subgradient is introduced.

$$(\nabla_{\mathbf{x}}\|\mathbf{x}\|_1)_u=\begin{cases}1, & \text{if } x_u>0\\ -1, & \text{if } x_u<0\\ [-1,1], & \text{if } x_u=0.\end{cases} \tag{41}$$

Puting this subgradient together with the gradient of the data fitting term, the uth gradient has the form

$$0=\begin{cases}x_u-y_u+\mu, & \text{if } x_u>0\\ x_u-y_u-\mu, & \text{if } x_u<0\\ [-y_u-\mu, y_u+\mu], & \text{if } x_u=0\end{cases} \tag{42}$$

Then, we have the optimum $\mathbf{x}^*$ with the uth entry

$$x_u=\begin{cases}y_u-\mu, & \text{if } x_u>0 \text{ or } y_u\geq\mu\\ y_u+\mu, & \text{if } x_u<0 \text{ or } y_u\leq-\mu\\ 0, & \text{if } -\mu\leq y_u\geq\mu.\end{cases} \tag{43}$$

We also write above operator as, known as soft-thresholding operator

$$\mathbf{x}=S_\mu(\mathbf{y}):=(|\mathbf{y}|-\mu)_+\text{sign}(\mathbf{y}). \tag{44}$$

The famous ISTA algorithm is a method with soft-thresholding the gradient-descent step.

final notes

We introduce these for the sake of understanding what optimization is doing, why ℓ_1 norm is more involved than ℓ_2 norm. But for this research project, we still also suggest to use the well-known iterative algorithms, e.g., sgd, adam, coordinate descent. These algorithms are all well embedded in python packages or in other languages.