A Tutorial on Machine Learning for Spatiotemporal Data Modeling

Revision: Matrix and Tensor Factorization

Xinyu Chen

February 3, 2022

2.1 (Low-Rank Matrix Factorization). For any partially observed data matrix $\boldsymbol{Y} \in \mathbb{R}^{N \times T}$ with the observed index set Ω, a matrix factorization algorithm can decompose $\boldsymbol{Y}$ into lower dimensional factor matrices $\boldsymbol{W} \in \mathbb{R}^{R \times N}$, $\boldsymbol{X} \in \mathbb{R}^{R \times T}$, and its loss function can be written as

$$f = \frac{1}{2} \sum_{(i,t) \in \Omega} \left(y_{i,t} - \boldsymbol{w}_i^\top \boldsymbol{x}_t\right)^2 + \frac{\rho}{2} \left(\sum_{i=1}^{N} \|\boldsymbol{w}_i\|_2^2 + \sum_{t=1}^{T} \|\boldsymbol{x}_t\|_2^2 \right), \tag{1}$$

where $\boldsymbol{w}_i \in \mathbb{R}^R$ is the ith column of $\boldsymbol{W}$, and $\boldsymbol{x}_t \in \mathbb{R}^R$ is the tth column of $\boldsymbol{X}$. The symbol $\|\cdot\|_2$ denotes the ℓ_2-norm.

1. Obtain the partial derivative with respect to $\boldsymbol{w}_i$, i.e., $\frac{\partial f}{\partial \boldsymbol{w}_i}$. Let $\frac{\partial f}{\partial \boldsymbol{w}_i} = \boldsymbol{0}$, what is the solution to $\boldsymbol{w}_i$?

   In this case, with respect to $\boldsymbol{w}_i$, the partial derivative is given by

$$\frac{\partial f}{\partial \boldsymbol{w}_i} = -\sum_{t:(i,t) \in \Omega} \boldsymbol{x}_t \left(y_{i,t} - \boldsymbol{w}_i^\top \boldsymbol{x}_t\right) + \rho \boldsymbol{w}_i. \tag{2}$$

If $\frac{\partial f}{\partial \boldsymbol{w}_i} = \boldsymbol{0}$, then we have

$$\begin{aligned} -\sum_{t:(i,t)\in\Omega} \boldsymbol{x}_t \left(y_{i,t} - \boldsymbol{w}_i^\top \boldsymbol{x}_t\right) + \rho \boldsymbol{w}_i = \boldsymbol{0} \\ \Longrightarrow -\sum_{t:(i,t)\in\Omega} \boldsymbol{x}_t y_{i,t} + \left(\sum_{t:(i,t)\in\Omega} \boldsymbol{x}_t \boldsymbol{x}_t^\top + \rho \boldsymbol{I}_R\right) \boldsymbol{w}_i = \boldsymbol{0}. \end{aligned} \tag{3}$$

Thus,

$$\boldsymbol{w}_i = \left(\sum_{t:(i,t)\in\Omega} \boldsymbol{x}_t \boldsymbol{x}_t^\top + \rho \boldsymbol{I}_R\right)^{-1} \sum_{t:(i,t)\in\Omega} \boldsymbol{x}_t y_{i,t}. \tag{4}$$

2. Obtain the partial derivative with respect to $\boldsymbol{x}_t$, i.e., $\frac{\partial f}{\partial \boldsymbol{x}_t}$. Let $\frac{\partial f}{\partial \boldsymbol{x}_t} = \boldsymbol{0}$, what is the solution to $\boldsymbol{x}_t$?

   In this case, with respect to $\boldsymbol{x}_t$, the partial derivative is given by

$$\frac{\partial f}{\partial \boldsymbol{x}_t} = -\sum_{i:(i,t)\in\Omega} \boldsymbol{w}_i \left(y_{i,t} - \boldsymbol{w}_i^\top \boldsymbol{x}_t\right) + \rho \boldsymbol{x}_t. \tag{5}$$

   If $\frac{\partial f}{\partial \boldsymbol{x}_t} = \boldsymbol{0}$, then we have

$$\begin{aligned} -\sum_{i:(i,t)\in\Omega} \boldsymbol{w}_i \left(y_{i,t} - \boldsymbol{w}_i^\top \boldsymbol{x}_t\right) + \rho \boldsymbol{x}_t = \boldsymbol{0} \\ \Longrightarrow -\sum_{i:(i,t)\in\Omega} \boldsymbol{w}_i y_{i,t} + \left(\sum_{i:(i,t)\in\Omega} \boldsymbol{w}_i \boldsymbol{w}_i^\top + \rho \boldsymbol{I}_R\right) \boldsymbol{x}_t = \boldsymbol{0}. \end{aligned} \tag{6}$$

   Thus,

$$\boldsymbol{x}_t = \left(\sum_{i:(i,t)\in\Omega} \boldsymbol{w}_i \boldsymbol{w}_i^\top + \rho \boldsymbol{I}_R\right)^{-1} \sum_{i:(i,t)\in\Omega} \boldsymbol{w}_i y_{i,t}. \tag{7}$$

3. How to use Alternating Least Squares (ALS) method to solve the following optimization problem:

$$\min_{\boldsymbol{W},\boldsymbol{X}} \frac{1}{2} \sum_{(i,t)\in\Omega} \left(y_{i,t} - \boldsymbol{w}_i^\top \boldsymbol{x}_t\right)^2 + \frac{\rho}{2}\left(\sum_{i=1}^{N} \|\boldsymbol{w}_i\|_2^2 + \sum_{t=1}^{T} \|\boldsymbol{x}_t\|_2^2\right). \tag{8}$$

- Initialize $\boldsymbol{W}$ and $\boldsymbol{X}$.
- Repeat
  - For $i = 1$ to N:
    - Update $\boldsymbol{w}_i$ by Eq. (4).
  - For $t = 1$ to T:
    - Update $\boldsymbol{x}_t$ by Eq. (7).
- Return $\boldsymbol{W}$ and $\boldsymbol{X}$.

2.2