

A Tutorial on Machine Learning for Spatiotemporal Data Modeling

Revision: Matrix and Tensor Factorization

Xinyu Chen

February 4, 2022

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

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

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

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

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

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

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

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

Thus,

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

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

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

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

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

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

Thus,

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

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

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

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

2.2 (An Alternating Direction Algorithm for Matrix Completion with Non-negative Factors).

Definition 1 (Nonnegative Matrix Factorization/Completion (NMFC))

For any partially observed and nonnegative data matrix $\mathbf{Y} \in \mathbb{R}_+^{N \times T}$ with the observed index set Ω , find lower dimensional nonnegative factor matrices $\mathbf{W} \in \mathbb{R}^{R \times N}$, $\mathbf{X} \in \mathbb{R}^{R \times T}$ such that $\sum_{(i,t) \in \Omega} (y_{i,t} - \mathbf{w}_i^\top \mathbf{x}_t)^2$ is minimized.

1. Write down the formula of this optimization problem.