Relationship between Matrix/Tensor Factorization and Linear Regression

Mengyan Zhang

September 3, 2018

Abstract

Linear regression aims to model the relationship between two variables by fitting a linear equation to observed data. Matrix or tensor factorization is to factorize matrix or tensor into two or more smaller size matrices or tensors, which can be viewed as a regression problem, where the independent variable is each entry in the matrix/tensor, the corresponding dependent variable is the estimated entry. Therefore, the methods for linear regression can also be applied to the matrix/tensor factorization. This tutorial aims to present these relationships between matrix/tensor factorization and linear regression in detail.

	Least Squares	**Maximum Likelihood**	**Maximum A Posterior**	**Bayesian Approaches**
LR	$\arg\min_{\boldsymbol{\theta}} \lVert \boldsymbol{y} - X\boldsymbol{\theta} \rVert_F^2$	$\arg\max_{\boldsymbol{\theta}} p(\boldsymbol{y} \mid X, \boldsymbol{\theta}, \sigma)$	$\arg\max_{\boldsymbol{\theta}} p(\boldsymbol{\theta} \mid X, \boldsymbol{y}, \sigma, \sigma_\theta)$	$p(\boldsymbol{\theta} \mid X, \boldsymbol{y}, \sigma, \sigma_\theta)$
MF	$\arg\min_{U,V} \lVert A - UV \rVert_F^2$	$\arg\max_{U,V} p(A \mid U, V, \sigma)$	$\arg\max_{U,V} p(U, V \mid A, \sigma, \sigma_U, \sigma_V)$	$p(U, V \mid A, \sigma, \sigma_U, \sigma_V)$
TF	$\arg\min_{E,R} \sum_{k=1}^{K} \lVert T_{:k:} - ER_kE^T \rVert_F^2$	$\arg\max_{E,R} p(T \mid E, R, \sigma)$	$\arg\max_{E,R} p(E, R \mid T, \sigma, \sigma_E, \sigma_R)$	$p(E, R \mid T, \sigma, \sigma_E, \sigma_R)$

Table 1: Relation between linear regression and matrix/tensor factorization. LR: linear regression; MF: matrix factorization; TF: tensor factorization

1 Linear Regression

For linear regression, given D dimensional inputs $\boldsymbol{x} \in \mathbb{R}^D$ and corresponding target $y \in \mathbb{R}$, we aim to find a linear model which can estimate the target value for unseen input accurately. Then linear model can be built as

$$\hat{y} = f(\boldsymbol{x}) + \epsilon = \boldsymbol{x}^T\boldsymbol{\theta} + \epsilon, \tag{1}$$

where $\boldsymbol{\theta} \in \mathbb{R}^D$ are the parameters we seek, and $\epsilon \sim \mathbb{N}(0, \sigma)$ is i.i.d. Gaussian observation noise.

1.1 Least Squares

Intuitively, we want to find the parameters which can estimate the target value as accurate as possible, i.e. the sum of squares of the vertical deviations as small as possible, which is known as Least squares. For N inputs $\boldsymbol{x_n} \in \mathcal{R}^D$ and corresponding target $y_n \in \mathbb{R}$, n = 1, ... , N, we find parameters $\boldsymbol{\theta}$ by,

$$\arg\min_{\boldsymbol{\theta}} \sum_{n=1}^{N} (y_n - \hat{y_n})^2, \tag{2}$$

where $\hat{y_n}$ is defined as Equation 1. In matrix form, we have

$$\underset{\boldsymbol{\theta}}{\arg\min} \quad \|\boldsymbol{y} - X\boldsymbol{\theta} - \sigma\|_F^2, \tag{3}$$

where $X := [\boldsymbol{x_1, ..., x_N}]^T \in \mathbb{R}^{N\times D}$ and $\boldsymbol{y} := [y_1, ..., y_N]^T \in \mathbb{R}^N$.

2 Matrix Factorization

Matrix factorization is to factorize of a matrix into a product of matrices. For a given matrix $A \in \mathbb{R}^{N\times M}$, it can be factorized as the product as the product of $U \in \mathbb{R}^{N\times D}, V \in \mathbb{R}^{D\times M}$, i.e.

$$\hat{A} = UV, \tag{4}$$

2.1 Least Squares

Similar as linear regression, we want the estimated matrix $\hat{A}$ (i.e. the product of factorized matrices U and V) as much similar as the original matrix A. Thus the simplest way is to minimize the sum of error squares,

$$\underset{U,V}{\arg\min} \quad \sum_{i=1}^{N}\sum_{j=1}^{M}(a_{ij} - \hat{a_{ij}})^2, \tag{5}$$

where $\hat{a_{ij}} = \boldsymbol{u_i^T v_j}$ is the (i,j) entry of the estimated matrix $\hat{A}$. In matrix form,

$$\underset{U,V}{\arg\min} \quad \|A - UV\|_F^2, \tag{6}$$

3 Tensor Factorization

More generally, we need to deal with random dimension tensors rather than 2-d matrix. Using a three-way tensor $T \in \mathbb{R}^{N\times K\times N}$ as example, we simplify the tensor factorization into multiple matrix factorizations, i.e. for each slice $T_{:k:}$, we have

$$\hat{T_{:k:}} = ER_kE^T, \quad \text{for} \quad k = 1, ..., K, \tag{7}$$

where $E \in \mathbb{R}^{N\times D}$, $R_k \in \mathbb{R}^{D\times D}$ are factorized matrices for slice k. Note that E will be the same for all slices.

3.1 Least Squares

Similar as matrix factorization, we still want to find the most similar estimated matrix by minimizing the mean squared error, i.e.

$$\underset{E,R}{\arg\min} \sum_{k=1}^{K} \|T_{:k:} - ER_kE^T\|_F^2, \tag{8}$$