SVD

May 3, 2018

Abstract

1 Eigenvalue Decomposition

Suppose there is $m \times m$ full rank symmetric matrix A, it has m different eigenvalue, and the eigenvalue is λ_i, and the corresponding uint eigenvector is x_i, then

$$\begin{aligned} Ax_1 &= \lambda_1 x_1 \\ Ax_2 &= \lambda_2 x_2 \\ &\cdots \\ Ax_m &= \lambda_m x_m \end{aligned} \tag{1}$$

then :

$$AU = U\Lambda \tag{2}$$

$$U = [x_1 \quad x_2 \cdots x_m] \tag{3}$$

$$\Lambda = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_m \end{bmatrix} \tag{4}$$

Since the symmetric matrix eigenvectors are orthogonal to each other, U is a orthogonal matrix, and the inverse matrix of the orthogonal matrix is eaqual to its transpose.

$$A = U\Lambda U^{-1} = U\Lambda U^T \tag{5}$$

2 Singular Value Decomposition

Suppose there is $m \times n$ matrix A, the aim is to find a group of orthogonal basis in n-dimensional space which is also orthogonal after A-transforms. If the orthogonal basis v is the eigenvector of the matrix $A^T A$, then the elements in v are orthogonal to each other because the matrix $A^T A$ is symmetric matrix.

$$
\begin{aligned}
{v_i}^T A^T A v_j &= {v_i}^T \lambda_j v_j \\
&= \lambda_j {v_i}^T v_j \\
&= \lambda_j v_i v_j = 0
\end{aligned} \tag{6}
$$

Then normalize the orthogonal basis after mapping:
Tips: $\begin{bmatrix} a \\ b \end{bmatrix} \cdot \begin{bmatrix} c \\ d \end{bmatrix} = [ac + bd] = \begin{bmatrix} a \\ b \end{bmatrix}^T \begin{bmatrix} c \\ d \end{bmatrix}$

$$
\begin{aligned}
Av_i \cdot Av_i &= (Av_i)^T A v_i \\
&= {v_i}^T A^T A v_i \\
&= \lambda_i v_i \cdot v_i = \lambda_i
\end{aligned} \tag{7}
$$

$\Rightarrow$

$$
|Av_i|^2 = \lambda_i \geq 0 \tag{8}
$$

So the unit vector is:

$$
u_i = \frac{Av_i}{|Av_i|} = \frac{1}{\sqrt{\lambda_i}} A v_i \tag{9}
$$

$\Rightarrow$

$$
Av_i = \sigma_i u_i \tag{10}
$$

σ_i is singular value, and $\sigma_i = \sqrt{\lambda_i}, 0 \leq i \leq k, k = Rank(A)$
Now expand the orthogonal vector $\{u_1, u_2, \cdots, u_k\}$ to $\{u_1, u_2, \cdots, u_m\}$ as a group of orthogonal vector in m-dimensional space. Meanwhile, expand $\{v_1, v_2, \cdots, v_k$ to $\{v_1, v_2, \cdots, v_n$ as a group of orthogonal vector and $\{v_{k+1}, v_{k+2}, \cdots, v_n$ in $Ax = 0$ solution space, when $i > k$, $\sigma_i = 0$

$$
\begin{aligned}
&A[v_1 v_2 \cdots v_k | v_{k+1} \cdots v_n] \\
= [u_1 u_2 \cdots u_k | u_{k+1} \cdots u_m] &\begin{bmatrix} \sigma_1 & & \\ & \ddots & 0 \\ & & \sigma k \\ & 0 & 0 \end{bmatrix}
\end{aligned} \tag{11}
$$

$$
A = \begin{bmatrix} u_1 \cdots u_k \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_k \end{bmatrix} \begin{bmatrix} {v_1}^T \\ \vdots \\ {v_k}^T \\ {v_{k+1}}^T \\ \vdots \\ {v_n}^T \end{bmatrix} \tag{12}
$$

Use the multiplication of matrix:

$$
A = \begin{bmatrix} u_1 \cdots u_k \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_k \end{bmatrix} \begin{bmatrix} {v_1}^T \\ \vdots \\ {v_k}^T \end{bmatrix} + \begin{bmatrix} u_{k+1} \cdots u_m \end{bmatrix} \begin{bmatrix} 0 \end{bmatrix} \begin{bmatrix} {v_{k+1}}^T \\ \vdots \\ {v_n}^T \end{bmatrix} \tag{13}
$$

Then matrix A can be decomposed as:

$$A = \begin{bmatrix} u_1 \cdots u_k \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_k \end{bmatrix} \begin{bmatrix} {v_1}^T \\ \vdots \\ {v_k}^T \end{bmatrix} \tag{14}$$

$$X = \begin{bmatrix} u_1 \cdots u_k \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_k \end{bmatrix} = \begin{bmatrix} \sigma_1 u_1 \cdots \sigma_k u_k \end{bmatrix} \tag{15}$$

$$Y = \begin{bmatrix} {v_1}^T \\ \vdots \\ {v_k}^T \end{bmatrix} \tag{16}$$

Then $A = XY$ is the full rank decomposition of matrix A.

3 SVD in Recommender System

The data matrix of *user and item* is huge. And there are many blanks in this matrix, so it is also an extrmely sparse matrix. We defined this scoring matrix as $R_{U \times I}$

Table 1: user/item matrix

user/item	1	2	3	4	$\cdots$
1	5	4	4.5	?	$\cdots$
2	?	4.5	?	4.5	$\cdots$
3	4.5	?	4.4	4	$\cdots$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$

The scoring matrix $R_{U \times I}$ can be decompesed as two matrices$P$ and Q:

$$R_{U \times I} = P_{U \times k} Q_{k \times I} \tag{17}$$

U is the number of users, I is the number of items, k is the rank of matrix R. Assuming that the known score is r_{ui}, the error between the true value and the predicted value is:

$$e_{ui} = r_{ui} - \widehat{r}_{ui} \tag{18}$$

The total error squared sum is:

$$SSE = \sum_{u,i} {e_{ui}}^2 = \sum_{u,i} (r_{ui} - \sum_{k=1}^{K} p_{uk} q_{ki}) \tag{19}$$

3.1 Basic SVD with Gradient Descent

$$\begin{aligned}
\frac{\partial}{\partial p_{uk}} SSE &= \frac{\partial}{\partial p_{uk}} \sum_{u,i} (e_{ui})^2 \\
&= 2e_{ui} \frac{\partial}{\partial p_{uk}} e_{ui} \\
&= 2e_{ui} \frac{\partial}{\partial p_{uk}} (r_{ui} - \sum_{k=1}^{K} p_{uk} q_{ki}) \\
&= 2e_{ui} q_{ki}
\end{aligned} \tag{20}$$

Explanation for the equation:
there is no summary symbol in the second step because only the equation which has u has the derivative result, other equations' derivative results are zero.

$$\begin{aligned}
p_{uk} &:= p_{uk} - \eta(-e_{ui} q_{ki}) = p_{uk} + \eta e_{ui} q_{ki} \\
q_{ki} &:= q_{ki} - \eta(-e_{ui} p_{uk}) = q_{ki} + \eta e_{ui} p_{uk}
\end{aligned} \tag{21}$$

There are two different ways to update the two arguments:

- **Batch Gradient Descent Algorithm**
  Update p,q after calculating all the predictions of known scores
- **Random Gradient Descent Algorithm**
  Update p,q after calculating one e_{ui}

We choose random gradient descent algorithm because we have to minimize the equation SSE, so we have to search in the direction of negative gradient, this may cause the gradient descent stop in the local optimal solution if we use bath gradient descent algorithm.