notes on SVD and its applications

Jian Wang

August 9, 2018

Abstract

Singular value decomposition (SVD) has a rich applications.

1 definition

The transpose ^{T} symbol can be generalized to hermitian conjugate †, but for simplicity, I will talk about real matrices and use transpose only.

$$A_{m\times n} = U\Sigma_{m\times n}V^T$$

Convention: Capital are for matrices, lower case letters are for column vectors.

$$U_{m\times m} = [u_1, u_2, \cdots, u_m]$$

each u_i is an $m \times 1$ column vector.

Similar for V

$$V_{n\times n} = [v_1, v_2, \cdots, v_n]$$

These either U or V sets are complete orthogonal unit vectors, in there m or n dimensional space.

Σ is diagonal semi-positive defined matrix, although it might not be square matrix. It can be represented by a set of singular values:

$$\Sigma \sim \mathrm{Diag}\big[\sigma_1, \sigma_2, \cdots, \sigma_{\min\{m,n\}}\big]$$

The convection is in descending order: $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_{\min\{m,n\}} \geqslant 0$

1.1 relation with eigen-decomposition

There are two symmetric matrices A^TA and AA^T

$$A^TA = V(\Sigma^T\Sigma)V^T \tag{1}$$

$$AA^T = U(\Sigma\Sigma^T)U^T \tag{2}$$

The eigenvalue is square of the singular value:

$$\Lambda \sim \mathrm{Diag}\big[\sigma_1^2, \sigma_2^2, \cdots, \sigma_{\min\{m,n\}}^2, 0, \cdots, 0_{\max\{m,n\}}\big]$$

If $m \neq n$, the large dimension matrix is always singular (not fully rank).

1.2 geometric interpretation

linear transformation = rotation $\times$ scaling $\times$ rotation

From this geometric interpretation, we can also see why V^T is defined with transpose in SVD. This is because if we multiply ΣV^T with right eigenvector v_i, v_i will be rotate to e_i direction and scaled by σ_i

The scaling is in orthogonal directions.

Then by multiplying U, e_i is mapped to u_i direction.

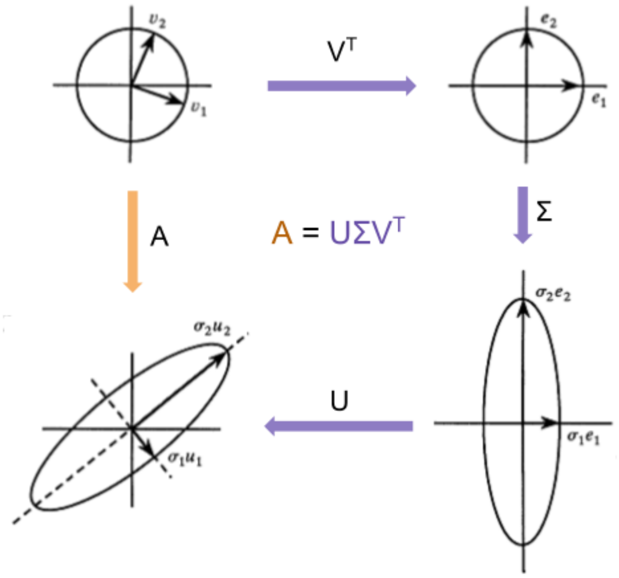

Figure 1: rotation scaling rotation. picture taken from the web: https://kevinbinz.com/2017/03/07/svd/

The scaling matrix $\Sigma_{m\times n}$ need some careful treatment.

1. $m = n$, all $\sigma_i > 0$: good
2. $m = n$, some $\sigma_i = 0$: compress in that direction
3. $m < n$, all $\sigma_i > 0$: naturally compressed
   these unlucky $n - m$ dimensions is just the extra columns of v_i
4. $m > n$, all $\sigma_i > 0$: total space expanded, but the effective subspace is still n-dimensional

2 application

2.1 least square fitting

$$Ax = b$$

For general $A_{m\times n}$ rectangle matrix, it is not always guaranteed that there exist solution for x. Instead of solving $Ax = b$, we are going to minimize $||Ax - b||_2$, and there is always solution for such a minimization problem.

Since unitary transformation doesn't change the 2-norm. The problem is equivalent to minimizing $||\Sigma V^T x - U^T b||_2$

Replacing with $y = V^T x$ and $c = U^T b$ without losing dimensionality, we are ending with a problem of minimizing $||\Sigma y - c||_2$

Three basic cases:

- Rank$(\Sigma) >$ Dim(c) : non-unique y and min=0
- Rank$(\Sigma) =$ Dim(c) : unique y and min=0.
- Rank$(\Sigma) <$ Dim(c) : unique y and min ≥ 0

Rank is just the numbers non-zero elements in Σ. The cause might be the dimension or the $\sigma = 0$ terms.

$$y_i = c_i/\sigma_i$$

And the formal solution looks like:

$$x = A^{-1}b = V\Sigma^+ U^T b$$

There are subtitles and clever cut-off regarding the zeros in Σ. See link svd and pseudo-inverse

2.2 principle component analysis

$$A = U\Sigma V^T = u_1\sigma_1 v_1^T + \dots$$

2.3 correlation matrix

Suppose $X_{n\times p}$ is a data set, n is the size of the sample, p is the parameters, features, etc..

For example, $X_{n\times p}$ comes from quantum Monte Carlo simulation.

Usually, it is enough to find the best value $\overline{X_p} = \frac{1}{n}\sum_{i=1}^{n} X_{ip}$ and their error $\sigma_p^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_{ip} - \overline{X_p})^2$, and sometimes we also need to check if the distribution shape of X_p is Gaussian.

However, one step further study assumes X_p are not independent, but still Gaussian. The multi-variable Gaussian distribution:

$$X_p \sim e^{-\frac{1}{2}(X_p-\mu_p)^T I (X_p-\mu_p)} \tag{3}$$

See this https://stats.stackexchange.com/questions/134282/relationship-between-svd-and-pca-how-to-use-svd-to-perform-pca

2.4 p-h symmetric Hamiltonian

$$H = \begin{pmatrix} 0 & M^T \\ M & 0 \end{pmatrix} \tag{4}$$

Hamiltonian like this has a special property — particle-hole symmetry:

If $\begin{pmatrix} x \\ y \end{pmatrix}$ is an eigenvector with eigenvalue λ, then $\begin{pmatrix} y \\ x \end{pmatrix}$ is also an eigenvector, but with eigenvalue $-\lambda$

There are many subtle things, regarding solving the system numerically. Firstly, the eigenvectors are redundant in half. Secondly, when λ close to zero, particles and holes might mixing.

The right thing to do is to solve the SVD of M, instead of diagonalizing H directly.

2.5 Schmidt decomposition of pure state

Two quantum system A and B, the big Hilbert space can use the basis of $|A_m\rangle|B_n\rangle$. The general pure state

$$|\Psi\rangle = \sum_{m\in A, n\in B} C_{mn}|A_m\rangle|B_n\rangle \tag{5}$$

There are $M \times N$ coefficients in C_{mn}

We can find new basis for the subspace A $|a_m\rangle$ and subspace B $|b_n\rangle$, such that a simpler decomposition exist:

$$|\Psi\rangle = \sum_{i\in\min\{M,N\}} c_i|a_i\rangle|b_i\rangle \tag{6}$$

3 the numeric methods

I don't know how it is calculated currently.