Program Study Note

Zhenzhao Tu

Nov 30 2021

1 The spectral theorem for symmetric matrices

Suppose $A \in \mathbb{R}^{n \times n}$ is symmetric. Then

1. every eigenvalue λ of A is a real number and there exists a (real) eigenvector $u \in \mathbb{R}^n$ corresponding to λ: $Au = \lambda u$;
2. eigenvectors corresponding to distinct eigenvalues are necessarily orthogonal:
$$Au^{(1)} = \lambda_1 u^{(1)}, Au^{(2)} = \lambda_2 u^{(2)}, \lambda_1 \neq \lambda_2, u^{(1)} u^{(2)} = 0.$$
3. there exists a diagonal matrix $D \in \mathrm{R}^{n \times n}$ and an orthogonal matrix $U \in \mathrm{R}^{n \times n}$ such that $A = UDU^T$. The diagonal entries of D are the eigenvalues of A and the columns of U are the corresponding eigenvectors:
$$D = diag(\lambda_1, \lambda_2, \lambda_3, ..., \lambda_n),$$
$$U = [u^{(1)}|u^{(1)}|u^{(2)}|...|u^{(n)}],$$
$$Au^{(n)} = \lambda_n u^{(n)}, i = 1, 2, 3, ..., n.$$

An orthogonal matrix U satisfies, by definition, $U^T = U^{-1}$, which means that the columns of U are orthonormal (that is, any two of them are orthogonal and each has norm one). The expression $A = UDU^T$ of a symmetric matrix in terms of its eigenvalues and eigenvectors is referred to as the **spectral decomposition** of A.
The spectral theorem implies that there is a change of variables which transforms A into a diagonal matrix.

Don't go into the proof yet. First state the SVD. Second, discuss some of its consequences at a high level. Third, prove it.

2 Singular Value Decomposition(SVD)

2.1 Definition of SVD(Singular Value Decomposition)

First set : $A \in \mathbb{R}^{m\times n}$, assume $n \leq m$, $r = rank(A) \leq n$
Then : $\exists\ U \in \mathbb{R}^{m\times r}$, $V \in \mathbb{R}^{r\times n}$ (Here U and V are both have orthogonal columns), and $\exists\ \Sigma \in \mathbb{R}^{r\times r}$.
That means Σ has to be a **diagonal** matrix with strictly positive entries such that:

$$A = U\Sigma V^T \tag{1}$$

Which express by:

$$A = \sum_{j=1}^{r} \sigma_j u_j v_j^T \tag{2}$$

where:

- U is an $m \times m$ orthogonal matrix.
- V is an $n \times n$ orthogonal matrix.
- Σ is an $m \times n$ matrix whose i^{th} diagonal entry equals the i^{th} singular value σ_i for i = 1,...,r. All other entries of Σ are zero.

2.2 Find a SVD

Let A be an $m \times n$ matrix with $\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_1 \geq 0$, and let r denote the number of nonzero singular values A.

Let $v_1, ..., v_n$ be an orthogonal basis of $\mathbb{R}^{m\times n}$, where v_i is an eigenvector of $A^T A$ with eigenvalue σ_i^2.

Theorem 2.1.1 *Let A be $m \times n$ matrix. Then A has a (not unique)singular value decomposition $A = U\Sigma V^T$, where U and V are as follows:*

- The columns of V are orthogonal eigenvectors $v_1, ..., v_n$ of $A^T A$ where $A^T Av_i = \sigma_i^2 v_i$.
- If $i \geq r$, so that $\sigma_i \neq 0$, then the i^th column of U is $\sigma_i^{-1} Av_i$. these columns are orthogonal, and the remaining columns of U are obtained by arbitrarily extending to an orthogonal basis for $\mathbb{R}^m$.

Theorem 2.1.1 Find a SVD of

$$A = \begin{pmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{pmatrix}$$

Step 1. Since A is not a symmetric matrix, it cannot have eigenvalue, we need to find the eigenvalue of A^TA. By computing in python, we can get A^TA easily

$$A^TA = \begin{pmatrix} 80 & 100 & 40 \\ 100 & 170 & 140 \\ 40 & 140 & 200 \end{pmatrix}.$$

We can compute three eigenvalues of A^TA: $\lambda_1 = 360, \lambda_2 = 90, \lambda_3 = 0$. By definition of singular values, we can find $\sigma_1 = \sqrt{360}, \sigma_2 = \sqrt{90}, \sigma_3 = \sqrt{0}$. By definition of Σ in SVD, it has to be a 2×3 matrix

$$\Sigma = \begin{pmatrix} \sqrt{360} & 0 & 0 \\ 0 & \sqrt{90} & 0 \end{pmatrix}.$$

Step 2. After found singular value, we need to find matrix V. We know that V is consist of orthogonal basis of eigenvectors of A^TA

$$v_1 = \begin{pmatrix} -1/3 \\ -2/3 \\ -2/3 \end{pmatrix}, v_2 = \begin{pmatrix} 2/3 \\ -1/3 \\ 2/3 \end{pmatrix}, v_3 = \begin{pmatrix} 2/3 \\ -2/3 \\ 1/3 \end{pmatrix}$$

These three vectors are columns of V

$$V = \begin{pmatrix} -1/3 & -2/3 & 2/3 \\ -2/3 & -1/3 & -2/3 \\ -2/3 & 2/3 & 1/3 \end{pmatrix}.$$

Step 2. Now, it is time to find the last one matrix U. There are two ways to find U. The first one is using definition $u_j = \frac{Av_j}{\sigma_j}$ for every j in $1, ..., n$. From this equation, we can simply write down

$$\sigma_1^{-1}Av_1 = \frac{1}{\sqrt{360}} \begin{pmatrix} 18 \\ 6 \end{pmatrix}, \sigma_2^{-1}Av_2 = \frac{1}{\sqrt{90}} \begin{pmatrix} 3 \\ 9 \end{pmatrix}$$

Since u_1 and $u - 2$ are columns of U and U is a strictly 2×2 matrix, we can write U as

$$U = \begin{pmatrix} \frac{3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} & \frac{-3}{\sqrt{10}} \end{pmatrix}.$$

In conclusion, we now have singular value decomposition

$$A = \begin{pmatrix} \frac{3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} & \frac{-3}{\sqrt{10}} \end{pmatrix} \begin{pmatrix} \sqrt{360} & 0 & 0 \\ 0 & \sqrt{90} & 0 \end{pmatrix} \begin{pmatrix} -1/3 & -2/3 & 2/3 \\ -2/3 & -1/3 & -2/3 \\ -2/3 & 2/3 & 1/3 \end{pmatrix}^T.$$

2.3 Proof

At first, let's set:

$$K = A^T A (\geq 0) \tag{3}$$

Then, K can be multiply by x^T and x:

$$x^T K x \tag{4}$$

By doing a small calculation, we get:

$$x^T A^T A x = \langle Ax, Ax \rangle \tag{5}$$

From equation (5), we can simply know length of Ax are equal and greater than 0:

$$||Ax||_2^2 \geq 0 \tag{6}$$

By using **Spectral Theorem**:

$$K = V \Lambda V^T \in \mathbb{R}^{n \times n} \tag{7}$$

The Λ in K is diagonal matrix, where V is an orthonormal matrix whose columns are the eigenvectors of $A^T A$ and where $r \leq n$ and $r = rank(A) = rank(A^T A)$. Now we define a quantity σ_i (the *singular value*) such this equation with λ_i: Now let's define an equation:

$$\sigma_j^2 = \lambda_j \quad (\sigma_j = \sqrt{\lambda_j} \geq 0) \tag{8}$$

Assuming there are λ_1 to λ_n, and each λ is greater than or equal to 0. The order of λ_j: $\lambda_j \geq \lambda_{j+1}$, $\forall\ j = 1, ..., (n-1)$. For the i-th eigenvector-eigenvalue pair, we have

$$A^T A v_i = (\sigma_i)^2 v_i.$$

It is easy to know that v_i is the eigenvector of $A^T A$. For now, assume that we have a full-rank matrix ($\sigma_i \geq 0$ for all i). Define a new matrix U is an orthonormal matrixr whose columns u_i such that

$$u_j = \frac{A v_j}{\sigma_j} \qquad u_j \in \mathbb{R}^m \text{and} v_j \in \mathbb{R}^n \tag{9}$$

Since u_i and v_i are orthonormal columns of U and V, u_j and v_j in these two equation satisfied $\langle u_j, u_k \rangle = \delta_{j,k}$ and $\langle v_j, v_k \rangle = \delta_{j,k}$, the δ in here is such this equation:

$$\delta_{j,k} = \begin{cases} 1, & j = k. \\ 0, & j \neq k. \end{cases} \tag{10}$$

We can easily know that K is a symmetric matrix, so we can write down it eigenvalue and eigenvector form:

$$Kv_j = \lambda_j v_j \tag{11}$$

Now, if we multiply A in both sides, we can prove a result below:

$$\begin{aligned} AKv_j &= \lambda_j Av_j \\ (AA^T)Av_j &= \lambda_j Av_j \end{aligned}$$

both sides divided by σ_j

$$(AA^T)u_j = \lambda_j u_j$$

From the last equation, we know two things: (1) u_j is an eigenvector of AA^T; (2) λ_j is an eigenvalue of AA^T .
Start at Av_j, we can re-write it as:

$$Av_j = \sigma_j u_j \tag{12}$$

To make our proof easier, here we define $j = 1...r$ ($assume$ **r=n**).
Don't forget U and V have orthogonal columns, they can be written in this form:

$$AV = U\Sigma \tag{13}$$

Since V is a $n \times n$ matrix with orthogonal columns, it is an orthogonal matrix. That means $V^T = V^{-1}$.
When we multiply V^{-1} in two sides, we can get our final answer:

$$A = U\Sigma V^T \tag{14}$$

3 The SVD for Derivatives and Integrals

I add this section is because it will help me to understand SVD in a new way. According to textbook, this is the clearest example of the SVD. At first we can think A as an *operator* instead of matrix. Now we can write down integral and derivative of function in terms of A and operator D:

$$Integral : Ax(s) = \int_0^s x(t)\, dt$$

$$Derivative : Dx(t) = \frac{dx}{dt}.$$

By the Fundamental Theorem of Calculus, D is the inverse of A. More exactly D is a left inverse of A. Derivative of integral equals original function, so $DA = I$. We call this as D ***is the pseudoinverse of*** A.