Lecture 3-Brushing Up Matrices

Multiplying and Factoring Matrices (Contd.)

Multiplying and Factoring Matrices

- Today's Discussion
- A=LU
- A=U$\boldsymbol{\Sigma}$V^T

Source: Section I.2 in Linear Algebra and Learning from Data (2019) by **Gilbert Strang**

Multiplying and Factoring Matrices

- Let's talk about elimination i.e., A=LU. It is solving like Ax=b which are row operations
- All those row operations that is expressed by L times U
- And it factors into lower triangular (L) times upper triangular (U)
- Take A (2x2) matrix, $A=\begin{bmatrix} 2 & 3 \\ 4 & 7 \end{bmatrix} \rightarrow \begin{bmatrix} \mathbf{2} & 3 \\ 0 & \mathbf{1} \end{bmatrix}$
- Multiplying row 1 by 2 and subtracting row 1 from row 2
- Upper triangular with the pivots **2** and **1** on the diagonal

Multiplying and Factoring Matrices

- Matrix A is L times U. L is the lower triangular matrix
- Take L= $\begin{bmatrix} \mathbf{1} & 0 \\ 0 & \mathbf{1} \end{bmatrix}$
- It has there the number that you used here. We multiplied 2 of the first row and subtracted it from 2nd row in A. So, we need a multiplier i.e., 2 there. So,
- L= $\begin{bmatrix} \mathbf{1} & 0 \\ 2 & \mathbf{1} \end{bmatrix}$
- Then it shows
- A= $\begin{bmatrix} \mathbf{1} & 0 \\ 2 & \mathbf{1} \end{bmatrix} \begin{bmatrix} \mathbf{2} & 3 \\ 0 & \mathbf{1} \end{bmatrix}$ =LU

Multiplying and Factoring Matrices

Parallel way to think of this 2 by 2 matrix A= $\begin{bmatrix} 2 & 3 \\ 4 & 7 \end{bmatrix}$ = $\begin{bmatrix} 2 & 3 \\ 4 & - \end{bmatrix}$ + $\begin{bmatrix} 0 & 0 \\ 0 & - \end{bmatrix}$ $\rightarrow$ $\begin{bmatrix} 2 & 3 \\ 4 & 6 \end{bmatrix}$ + $\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$ =L+U

Split it into first row and column in one piece

Elimination has taken the original matrix. It's split and these are both rank 1

So, L+U= [col l_1] [row u_1^T] + [col l_2] [row u_2^T]

Our idea is that it gives the breakdown. And this is, by column times row rule, that's LU

$A = \begin{bmatrix} & & \\ & & \\ & & \end{bmatrix}$ = (col1)(row1)+ $\begin{bmatrix} 0 & 0 & 0 \\ 0 & A22 & A23 \\ 0 & A32 & A33 \end{bmatrix}$ = (col1)(row1)+ (col2)(row2)+ $\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & A33 \end{bmatrix}$

Start with A, break it up into LU, where LU is the first column times row. And then the next pieces are the rest of the matrix

Second part, second column times the second row, maybe divide by the pivot to make it correct. And then A3 would be the rest of that matrix

Multiplying and Factoring Matrices

Singular Value Decomposition (SVD) $A=U\boldsymbol{\Sigma}V^T$
Compare with $S=Q\boldsymbol{\Lambda}Q^T$

For a rectangular matrix, the whole idea of eigenvalues is snapped because if Ax in n dimensions, output will come in m dimensions. $Ax = \boldsymbol{\lambda}x$ is not even possible if A is rectangular

A (rectangular matrix)=$U\boldsymbol{\Sigma}V^T$ ➞ U and V are singular vectors and $\boldsymbol{\Sigma}$ is singular values matrix of rank r, i.e.,

$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\sigma}_1 & & \\ & \boldsymbol{\sigma}_2 & \\ & & \boldsymbol{\sigma}_r \end{bmatrix}$ and other entries are zeros

Now A^TA =[mxn rows][nxm columns]=(nxn) matrix= Symmetric matrix. Since eigen values are ≥ 0, A^TA is positive definite (p.d) matrix

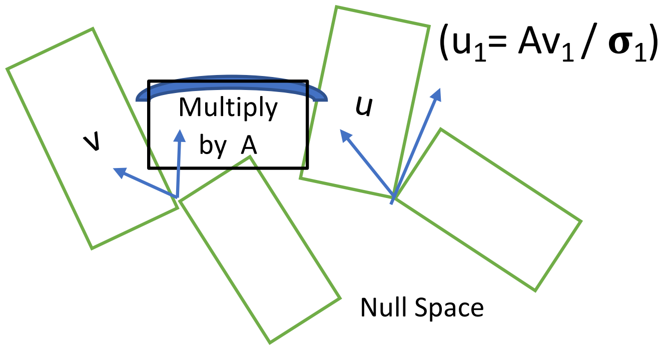
Multiplying and Factoring Matrices

Matrix can be factored $A^TA= V\Lambda V^T$ where V is orthogonal eigen vector matrix, V^T is its transpose and eigen value $\Lambda \geq 0$, positive definite

AA^T = (mxm) large matrix with same eigen values $\Lambda = U\Lambda U^T$, U is eigen vectors. Use U's and V's. They are same.

Matrix A is rectangular and hasn't eigenvectors. Av is sigma, singular value ($\boldsymbol{\sigma}$) times u. That's the first entry and the second entry and the rth entry. Stop at r, the rank. So,

$$\left.\begin{array}{l} Av_1= \boldsymbol{\sigma}_1 u_1 \\ Av_2= \boldsymbol{\sigma}_2 u_2 \\ . \\ . \\ Av_r= \boldsymbol{\sigma}_r u_r \end{array}\right\} \qquad A\begin{bmatrix} v_1, v_2,\ldots v_r \end{bmatrix} = \begin{bmatrix} u_1, u_2,\ldots u_r \end{bmatrix} \begin{pmatrix} \boldsymbol{\sigma}_1 & & \\ & \boldsymbol{\sigma}_2 & \\ & & \boldsymbol{\sigma}_r \end{pmatrix}$$



$$\Rightarrow AV= \boldsymbol{\Sigma} U \Rightarrow A= U\boldsymbol{\Sigma} V^T$$

$$\left.\begin{array}{l} Av_{r+1}= 0 \\ .. \\ Av_n= 0 \end{array}\right\} \text{Null Space}$$

$A^TA=V\boldsymbol{\Sigma}^TU^TU\boldsymbol{\Sigma}V^T= V(\boldsymbol{\Sigma}^T\boldsymbol{\Sigma})\, V^T$ where, V=evector of A^TA and $\boldsymbol{\sigma}^2$ evalues of A^TA and A^TA is symmetric, positive definite

Multiplying and Factoring Matrices

Using A^TA expression in getting u's

$Av_1 / \boldsymbol{\sigma}_1 = u_1$ .. .. $Av_r / \boldsymbol{\sigma}_r = u_r$

We've chosen v's and $\boldsymbol{\sigma}$'s. For A^TA, eigenvectors are v's and eigenvalues are sigma squared. Now we want u

$u_1^Tu_2 = 0$ (u_1 and u_2 are orthogonal)

$u_1^Tu_2 \Rightarrow (Av_1 / \boldsymbol{\sigma}_1)^T (Av_2 / \boldsymbol{\sigma}_2) \Rightarrow (v_1 A^TAv_2 / \boldsymbol{\sigma}_1 \boldsymbol{\sigma}_2) \Rightarrow v_1^T \boldsymbol{\sigma}_2^2 v_2 / \boldsymbol{\sigma}_1 \boldsymbol{\sigma}_2$ (v_1 and v_2 are orthogonal and v_2 evector of A^TA)
$\Rightarrow (\boldsymbol{\sigma}_2 / \boldsymbol{\sigma}_1)\, v_1^T v_2 = 0$

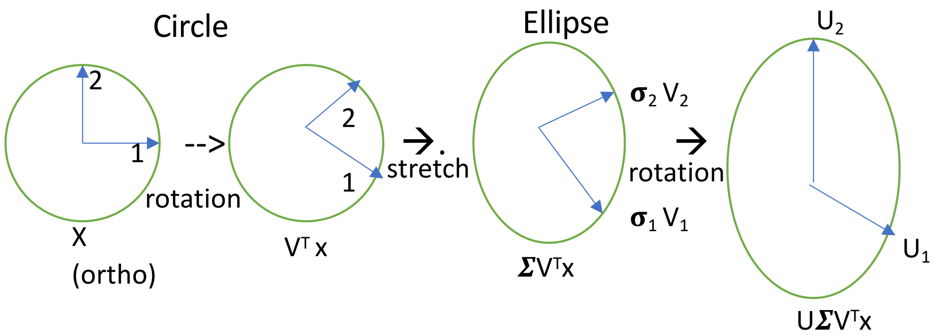
If we have a matrix A, say 5,000 by 10,000, why is it a mistake to use A^TA in the computation? It's very big and very expensive.

Actual computational methods are quite different

Because A^TA, it's symmetric, positive definite, we made the proof

Multiplying and Factoring Matrices

We take $A=U\Sigma V^T$ (U and V ortho. Matrices) Then $Ax=U\Sigma V^Tx$



So this SVD is telling us something quite remarkable that every linear transformation, every matrix multiplication factors into a rotation times a stretch times a different rotation, but possibly different.

Actually, when would the u be the same as a v?

A square. Because A would have to be square. And we want this to be the same as $Q\Lambda Q^T$ if they're the same. U's would be the same as the V's when the matrix is symmetric. And we need it to be positive definite as $\sigma_1 > \sigma_2 > \ldots . > \sigma_r > 0$

$A= U\Sigma V^T$ So $A=Q\Lambda Q^T$

Why is that? Because our convention is these σ's are greater or equal to 0. It's going to be the same, so far a positive definite symmetric matrix, the S that we started with is the same as the A. the Q is the U, the Q^T is the V transpose, then λ is the σ

Multiplying and Factoring Matrices

$A^T \quad A \quad = \quad V \Lambda V^T$ (V is orthogonal and $\Lambda \geq 0$)

$$\underbrace{A^T}_{n \times m} \ \underbrace{A}_{n \times m} = \underbrace{V \Lambda V^T}_{n \times n}$$

$AA^T = U \Lambda U^T$ (nxn)

$$A = \begin{pmatrix} 3 & 4 \\ 0 & 5 \end{pmatrix}$$

$\boldsymbol{\sigma}_1 \leq \boldsymbol{\lambda}_1 \leq \boldsymbol{\lambda}_2 \leq \boldsymbol{\sigma}_2$

$\boldsymbol{\sigma}_1\boldsymbol{\sigma}_2 = 15 = \boldsymbol{\sigma}_1\boldsymbol{\sigma}_2$

$$A = \underset{(m \times r)}{\begin{pmatrix} u_1 & u_2 & u_r \end{pmatrix}} \underset{(r \times r)}{\begin{pmatrix} \boldsymbol{\sigma}_1 & & \\ & \boldsymbol{\sigma}_2 & \\ & & \boldsymbol{\sigma}_r \end{pmatrix}} \underset{(r \times n)}{\begin{pmatrix} v_1^T \\ v_2^T \\ v_r^T \end{pmatrix}}$$

$$AA^T = \underset{(m \times m)}{\begin{pmatrix} u_1 & u_2 & u_m \end{pmatrix}} \underset{(m \times n)}{\begin{pmatrix} \boldsymbol{\sigma}_1 & & \\ & \boldsymbol{\sigma}_2 & \\ & & \boldsymbol{\sigma}_r \\ 0 & 0 \dots & 0 \end{pmatrix}} \underset{(n \times n)}{\begin{pmatrix} v_1^T \\ v_2^T \\ v_n^T \end{pmatrix}}$$

$= u_1 \boldsymbol{\sigma}_1 v_1^T =$ rank 1 matrix

That all these zeros when you multiply out, just give nothing, so that really the only thing that non-zero is in these bits. But there is a complete one.

So what are these extra u's that are in the null space of A, AA^T or A^TA? so two sizes, the large size and the small size. But then the things that count are all in there.

Multiplying and Factoring Matrices

Polar decomposition of a matrix. So a polar means-- what's polar in-- for a complex number i.e. $e^{i\theta r}$ and in the case of real number it is orthogonal of size 1 Every matrix factors into a symmetric matrix times an orthogonal matrix.

$A= U\Sigma V^T = SQ$

From SVD,

$(U\Sigma V^T)$ (UV^T)=SQ $(U\Sigma V^T$ → symmetric and (UV^T) → orthogonal

Key fact-- if we have a big matrix of data, A, and if I want to pull out of that matrix the important part, so that's what data science has to be doing

Out of a big matrix, some part of it is noise, some part of it is signal

We're looking for the most important part of the signal here. So we're looking for the most important part of the matrix.

Multiplying and Factoring Matrices

Well, let us look at $U\boldsymbol{\Sigma}V^T$

It's a rank one piece

So the simple matrix building block is like a rank one matrix, a something, something transpose

And what should I pull out of that as being the most important rank one matrix that's in that product?

That is, $u_1\boldsymbol{\sigma}_1 v_1^T$ = rank 1 matrix

Back-up Slide

Multiplying and Factoring Matrices

$AA^T = U\boldsymbol{\Sigma}V^TV\boldsymbol{\Sigma}^TV^T = U(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T)U^T$ (as $V^TV=I$) U is evectors of AA^T
A^TA and AA^T are two different. So each has its own eigen vectors and we use both. It's just perfect as far as it goes, but it hasn't gone to the end yet because we could have double eigenvalues and triple eigenvalues, and all those horrible possibilities.
Suppose a matrix has a symmetric matrix, has a double eigenvalue. Take an example

$$S= \begin{pmatrix} 1 & & \\ & 1 & \\ & & 5 \end{pmatrix} \begin{pmatrix} X \\ y \\ 0 \end{pmatrix} = \begin{pmatrix} X \\ Y \\ 0 \end{pmatrix}$$

What's the deal with eigenvectors for that matrix 1, 1, 5? So 5 has got an eigenvector. You can see what it is, 0, 0, 1 eigenvectors that go with $\lambda=1$ for that matrix? What would be eigenvectors for a $\lambda=1$? There was a whole plane of them. Any vector of the form x, y, 0, i.e., the whole plane of eigenvectors

And we need to pick two that are orthogonal, which we can do. And then they have to be-- in the SVD those two orthogonal vectors have to go to two orthogonal vectors. We conclude that the V's the singular vectors should be eigenvalues