Singular Value Decomposition

For a rectangular matrix, the whole idea of eigenvalues is shot because if we multiply A times a vector x in n dimensions, out will come something in m dimensions and it's not going to equal λx. So $Ax = \lambda x$ is not even possible if A is rectangular. And even if A is square, what are the problems, just thinking for a minute about eigenvalues? The case $S = Q\Lambda Q^T$ is the great case where I have a symmetric matrix and then it is got a full set of eigenvalues and eigenvectors and they are orthogonal, all good. But for a general square matrix, eigenvalues are complex or the eigenvectors are not orthogonal. So comes SVD.

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

There are two sets of singular vectors, not one. For eigenvectors, we just had one set, the $Q's$. Now we have a rectangular matrix, we have got one set of left eigenvectors in m dimensions, and we have got another set of right eigenvectors in n dimensions. And numbers in between are not eigenvalues, but singular values. Σ is a diagonal matrix with singular values on the diagonal. So again diagonal matrix is in the middle. But the numbers on the diagonal are all positive or 0.

So the first step we have to do, the math step is to show that any matrix can be factored into a form like $U\Sigma V^T$. That's the parallel to the spectral theorem that any symmetric matrix could be factored that way.

The key is that A^TA is a great matrix. It will be positive definite. It's eigenvalues are greater or equal to 0. And that will mean that we can take their square roots.

$$A^TA = V\Lambda V^T$$

And AA^T is a different guy, it will have a different shape. Again

$$AA^T = U\Lambda U^T$$

the heart of the thing, the non-zero eigenvalues are the same.

Looking for $Av = \sigma u$. this is what takes the place of $Ax = \lambda x$. And we stop for #rank times, because after that the $\sigma = 0$. In a word, we are looking for a bunch of orthogonal vectors v so that when we multiply then by A , we get a bunch of orthogonal vectors u.

We can also rewrite it into a matrix form

$$AV = U\Sigma$$

the columns of V, U are v, u respectively. Then we are going to put V^T on the right hand side

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

With the above factorization. Looking at A^TA again.

$$A^TA = V\Sigma^TU^TU\Sigma V^T = V\Sigma^T\Sigma V^T$$

What does it tell us?

$$\Sigma^T\Sigma = \Lambda$$

Same as AA^T.

Let us take the final step and established the SVD.

So the final step is to remember what we are going for $Av_i = \sigma_i u$. What we have to deal with now is we haven't quite finished. 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. And if we have triple eigenvalues or double eigenvalues, then what's the deal with eigenvectors if we have double eigenvalues? Suppose a matrix say a symmetric matrix has a double eigenvalue. We want to show $u_i^T u_k = 0$. Taking u_1, u_2 as an example.

$$u_1^T u_2 = (\frac{Av_1}{\sigma_1})(\frac{Av_2}{\sigma_2}) = \frac{v_1^T A^T A v_2}{\sigma_1 \sigma_2} = \frac{v_1^T {\sigma_2}^2 v_2}{\sigma_1 \sigma_2}$$

the last step is because v_2 is an eigenvector of $A^T A$. And v_1, v_2 is orthogonal. Hence $u_1^T u_2 = 0$.

We can also find

$$\sigma_1 \leq \lambda_1 \leq \lambda_2 \leq \sigma_2$$

$$\sigma_1 \sigma_2 = \lambda_1 \lambda_2$$

And the SVD tells us every matrix multiplication factors into a rotation times a stretch times a different rotation.

Now there comes another interesting question here. How to count the parameters in SVD factorization.

Suppose A is a 2 by 2 matrix. Actually we need four numbers. there are two numbers σ_1, σ_2 in Σ. And the other two matrices both have one parameters -- angle. And if we did 3 by 3 matrix, again there are three numbers in Σ , but how many in 3D rotation? It turns out to be three. Roll, pitch and yaw are somehow stands for the three parameters in reality. We can easily calculate for four dimension space.

In addition, they are two ways to represent $A = U\Sigma V^T$, the two orthogonal matrices can be a big one or a small one.

Let's do something nice at the end -- another factorization of A. That is famous in engineering, and it's famous in geometry -- **Polar decomposition**. All we will see is that it's virtually here.

$$A = SQ$$

We can get it quickly from SVD.

$$A = (U\Sigma U^T)(UV^T) = SQ$$

in mechanical engineering language tells us that any strain (like stretching of elastic thing) has a symmetric kind of stretch and a internal twist.

The idea to use all of then on a matrix of data. And we see the key fact. The key fact -- if we have a big matrix of data A, and if we want to pull out of that matrix the important part, so that is 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 are looking for the most important part of the signal here. In a way, the biggest numbers though we don't look at individual numbers. So what is the biggest part of the matrix? The answer is $u_1 \sigma_1 v_1^T$. It is the top principal part -- the biggest rank 1 matrix.