

PROFESSOR: The previous video was about positive definite matrices. This video is also linear algebra, a very interesting way to break up a matrix called the singular value decomposition. And everybody says SVD for singular value decomposition. And what is that factoring? What are the three pieces of the SVD?

So this is the fact is every matrix, rectangular, every matrix factors into-- these are the three pieces. $U \Sigma V^T$. People use those letters for the three factors.

The factor U is an orthogonal matrix, an orthogonal matrix. The factor Σ in the middle is a diagonal matrix. The factor V^T on the right is also an orthogonal matrix. So I have orthogonal, diagonal, orthogonal, or physically, rotation, stretching, rotation. Now we have seen three factors for a matrix, V, λ, V^{-1} . What's the difference? What's the difference between this SVD, this, and the $V, \lambda, V^T, V^{-1}, V, \lambda, V^{-1}$ for diagonalizing other matrices?

So the λ is diagonal and the Σ is diagonal, but they're different. The key point is I now have two different matrices, not just V and V^{-1} , but two different matrices. But the new great advantage is they are orthogonal matrices, both of them. So by going to-- and I can do it for rectangular matrices also. Eigenvalues really worked for square matrices.

Now we really are-- we have two. We have an input matrix and an output matrix. In those spaces m and n can have different dimensions. So by allowing two separate bases, we get rectangular matrices, and we get orthogonal factors with, again, a diagonal. And this is called-- these numbers Σ instead of eigenvalues, are called singular values. So these are the singular values. These are the singular vectors, the right singular vectors and the left singular vectors. That's the statement of the factorization.

But we have to think a little bit, what are those factors? What are the-- can we see why this works? So I want that. And let me do, as you see this coming, I'll look at $A^T A$. I like $A^T A$. So $A^T A$ will be, I transpose this. $V \Sigma^T U^T$, right? That's $A^T A$. Then I multiply by A $U \Sigma V^T$. And what do I have? Well, I've got six matrices.

But $U^T U$ in here is the identity, because U is an orthogonal matrix. So I really have just the V on one side, a $\Sigma^T \Sigma$, that'll be diagonal, and a V^T the

right. This I recognize. This I recognize. Here is a single V , a diagonal matrix, a V transpose. What I'm showing you here, what we reached is the eigenvalue, the diagonalization, the usual eigenvalues are in here and the eigenvectors are in here. But the matrix is $A^T A$.

Once again, A was rectangular and completely general and we couldn't see perfect results. But when we went to $A^T A$, that gave us a positive semidefinite matrix, symmetric for sure. Its eigenvectors will be orthogonal. That's how I know this V matrix, the eigenvectors for this symmetric matrix, are orthogonal and the eigenvalues are positive. And they're the squares of the singular value. So this is telling me the λ s for $A^T A$ are the σ^2 s for A . For A itself. λ is the same. λ for $A^T A$ is σ^2 for the matrix A .

Well that tells me V , that tells me σ , and U disappeared here because $U^T U$ was the identity. It just went away. How would I get hold of U ? Well, here's one way to see it. I multiply A times A^T in that order, in that order. So now I have $U \sigma V^T$ times the transpose, which is the $V \sigma^T U^T$ -- I'm having a lot of fun here with transposes.

But $V^T V$ is now the identity in the middle. So what do I learn here? I learn that U is the eigenvector matrix for AA^T . So these have the same eigenvalues, A times B has the same eigenvalues as B times A in this case, it comes out here. Same eigenvalues. This has eigenvectors V , this has eigenvectors U , and those are the V and the U in the singular value decomposition.

Well, I have to show you an example I have to show you an example and an application, and that's it. So here's an example. Suppose A , I'll make it a square matrix, $\begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix}$, not symmetric. Certainly not positive definite. I wouldn't use the word because that matrix is not symmetric. But it's got an SVD, three factors.

And I work them out. This is the orthogonal matrix. I have to divide by square root of 5 to make it unit vectors. Oops, that's not going to work. How about that? The two columns are orthogonal and that's a perfectly good U . And then in the σ , I got, well that's a-- oh, I did want 1 and 1. I did want 1 and 1, yes. So I have a singular matrix, determinant 0, singular matrix. So my eigenvalues will be 0 and it turns out square root of 10 is the other eigenvalue for that-- other singular value for this guy.

And now I'll put in the V transpose matrix, which is 1, 1, and 1, minus 1 is it? And those have length square root of 2, which I have to divide by. Well, I didn't do that so smoothly, but the result is clear. U , σ , V transpose, so here's the σ . And the singular values of this matrix are square root of 10 and then 0 because it's a singular matrix. And the eigenvectors, well the singular vectors of the matrix are the left singular vectors and the right singular vectors. That looks good to me.

And now the application to finish. A first application is, well, very important. All the time in this century, we're getting matrices with data in them. Maybe in life sciences, we test a bunch of sample people for genes. So I have a-- my data comes somehow-- I have a gene expression matrix. I have samples, people, people 1, 2, 3 in those columns. And I have in the rows, let me say four rows, I have genes, gene expressions.

That would be completely normal. A rectangular matrix, because the number of people and the number of genes is not the same. And in reality, those are both very, very big numbers, so I have a large matrix. And out of it, I want to-- and each number in the matrix is telling me how much the gene is expressed by that person. We may be searching for genes causing some disease. So we take several people, some well, some with the disease, we check on the genes. We get a big matrix and we look to understand something about of it.

What can we understand? What are we looking for? We're looking for the correlation, the connection, between some combination maybe of genes and some-- we're looking for a gene people connection here. But it's not going to be person number one. We're not looking for one person. We're going to find a mixture of those people, so we're going to have sort of an eigensample, eigenpeople. Oh, that's a terrible-- eigenperson would be better. So I think I'm seeing an eigenperson. Let me see where I'm going to put this.

So yeah, I think my matrix would be written-- oh, here is the main point. That just as I see in this example, it's the first vector and the first vector and the biggest σ that are all important. Well, in that example the other σ was 0, nothing. But in this example, I'll probably have three different σ s. But the largest σ , the first, the U_1 and the V_1 , it's that combination that I want. I want $U_1 \sigma_1 V_1^T$, the first eigenvector of A transpose A and of AA^T .

And the first singular, the biggest singular value, that's the information. That's the best sort of put together person, eigenperson, combination of these people and the best combination of

genes. It has the-- in statistics, I would say the greatest variance. In ordinary English, I would say the most information. The most information in this big matrix is in this very special matrix with only rank one, only a single column repeated. A single row repeated, and a number σ_1 , the number that tells me that. Because remember, U is a unit vector. V is a unit vector. It's that number σ_1 that's selling me.

So it's like that unit vector times that number, key number, times that unit vector, that's this. I'm talking here about principle component analysis. I'm looking for the principle component, this part. Principle component analysis. A big application in applied statistics. You know, in large scale drug tests, statisticians really have a central place here. And this is on the research side, to find the-- get the information out of a big sample.

So U_1 is sort of a combination of people. V_1 is a combination of genes. σ_1 is the biggest number I can get. So that's PCA, all coming from the singular value decomposition. Thank you.