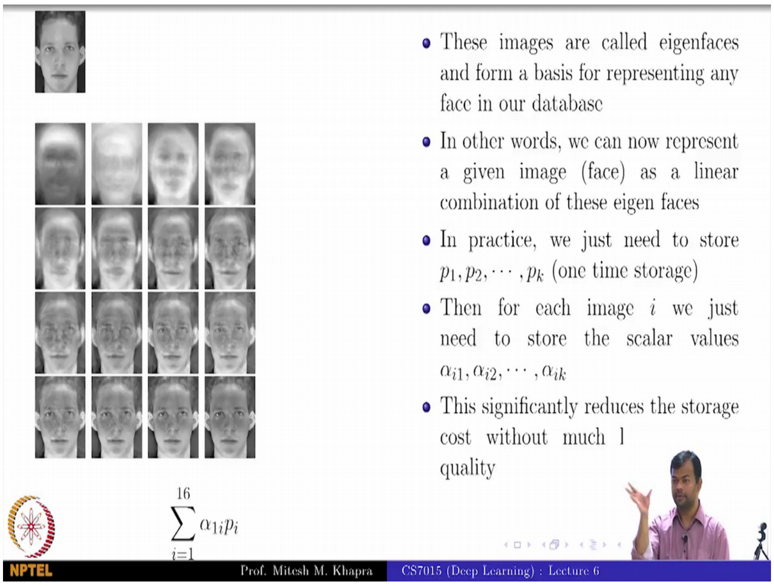
Deep Learning
Prof. Mithesh M. Khapra
Department of Computer Science and Engineering
Indian Institute of Technology, Madras

Module – 6.8
Lecture – 06
Singular Value Decomposition

(Refer Slide Time: 00:17)



So, with this we will move on to the last topic in the so, that is something that you will have to so, the way I would do it, right is that you keep aside some 100 images from your data as validation data, ok. Now once you have learned these eigenvectors, try to compute the reconstruction error for these 100 images. And just vary it, right 200,000, 10,000 written as many dimensions as you can, and see at what point is a reconstruction error, ok. For you, right and this is assuming that you have some notion of what is a reasonable reconstruction error. So, we all know that the minimum is 0, right.

But if you have 0.5, then maybe for face database it might be, ok. Right, but if it is a database where you are trying to look at mechanical parts, right. So, suppose you are looking at motors and rotors from a machine assembly. Now there you want to be able to distinguish minor detects defects on this and a detect could a defect could actually just be one single or 2 pixels getting different from the original image, right. So, there the reconstruction loss would be much needs to be much more robust, you get the point? So, it depends on your application. So, you will have to take some validation data either have a domain expert to tell you what is reasonable or go by the number that you get, right. And this is the validation error that I get.

So, everyone understands the question and perhaps the answers, ok. So, we now go to the last module.

Student: (Refer Time: 01:40).

Yeah, if you can.

Student: (Refer Time: 01:41).

Yes, you can now project any face into this database A. So, that is the Eigen basis that you have got, you have got the basis vectors, now any data you can project onto this basis.

Student: (Refer Time: 01:54).

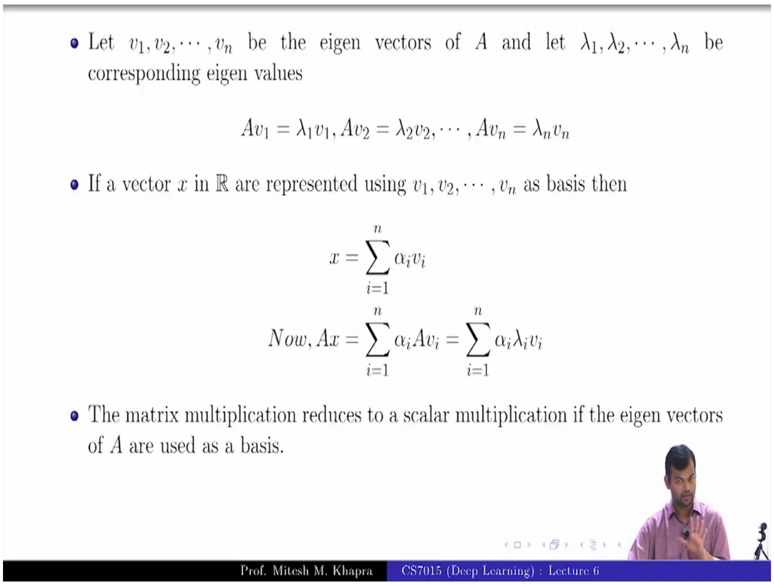
Now, so, if you are trying to learn these eigenvectors by say using 100 images all of which belonging to a particular demographic, say all Gaussian images, right. And now at the runtime you have an Asian image, then you will have; obviously, have some error right, but you have large even of data, say if you have if you are constructing this from million images, then it should generalize that is I mean just as for any machine learning algorithm, right.

The training it from small data and you bring out some outlier at test time it is not going to work, right, but if you have reasonable data it should generalize , any other questions? To calculate the eigenvectors x is m cross 10 m cross 10 k yes. So, so, so now, we move on to the last topic for the basic portion, and the next class we will do auto encoders will be back to dpl networks so, singular value decomposition, right.

(Refer Slide Time: 02:50)

So, this is actually the stuff that I need an important theorem from here at multiple 2 places in the course. So now, before doing the, right let us get some more perspective on what eigenvectors do and why are they actually important, ok.

(Refer Slide Time: 03:05)



So, let v 1 to v n be the eigenvectors of a and let lambda one be the corresponding Eigen values. So, we know this a v 1 equal to lambda v 1 and so on, ok. Now suppose all the vectors in R should be R raised to n, ok. So, if a vector x belonging to Rn can be represented using this basis, ok. Now what if I am interested in the operation A into x, what is the advantage of representing it using this basis? So, this is what you are saying the other day, right.

Student: (Refer Time: 03:54).

What is Ax it is a matrix vector multiplication, right. And it is going to be a heavy computation. Now if all my vectors in Rn are represented using the eigenvector as the basis, what happens to this matrix operation?

Student: (Refer Time: 04:16).

It reduces to.

Student: (Refer Time: 04:21).

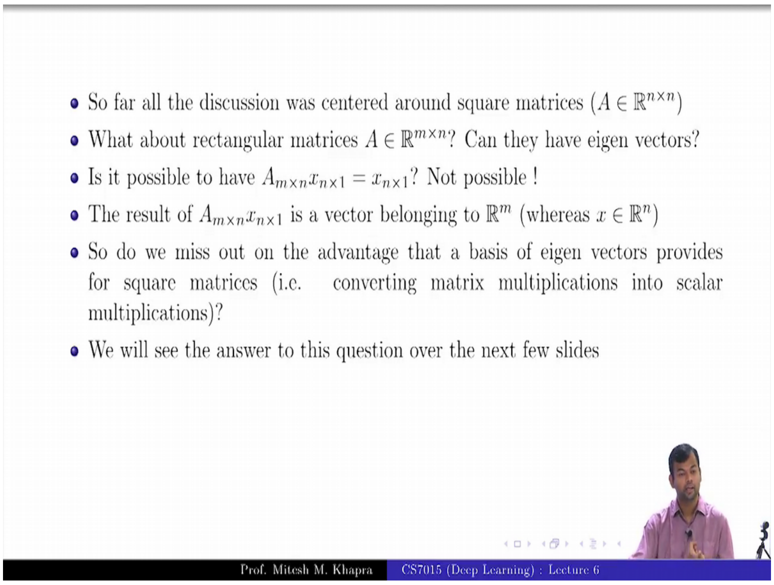
Let us see so, I was interested in Ax, but I know x is this. So, you get this step, and what happens finally? Do you have the matrix anywhere here. So, what happens to the matrix operation?

Student: (Refer Time: 04:41).

It reduces to a sum of scalar operations, right. If your vectors were representing using the eigenvector as a basis, ok.

So, this is one reason why this is important, right. So, you can now get away of the get rid of the matrix operations and just do scalar operations, right. Ok. So now, there is a catch here which I am going to ignore, just to try it if I bring in the catch you guys will get confused. So, I will ignore if anyone has a doubt maybe talk to me after the class, but for now let us go with the fact that the matrix operation reduces to a scalar operation, ok.

(Refer Slide Time: 05:19)



Now, so far what we have done is discussed square matrices, I have said that they are the villains of linear algebra, but who are the super villains of linear algebra , rectangular matrices everyone says that, but why. Imagine what they do to a vector yeah. So, can rectangular matrices have an eigenvector?

Student: (Refer Time: 05:45).

Yes; obviously, yes that I mean any matrix can have an eigenvector.

Student: No.

No why? Can you write something of this form? You can not, right. Because when the matrix operates on an n dimensional vector what does it give you?

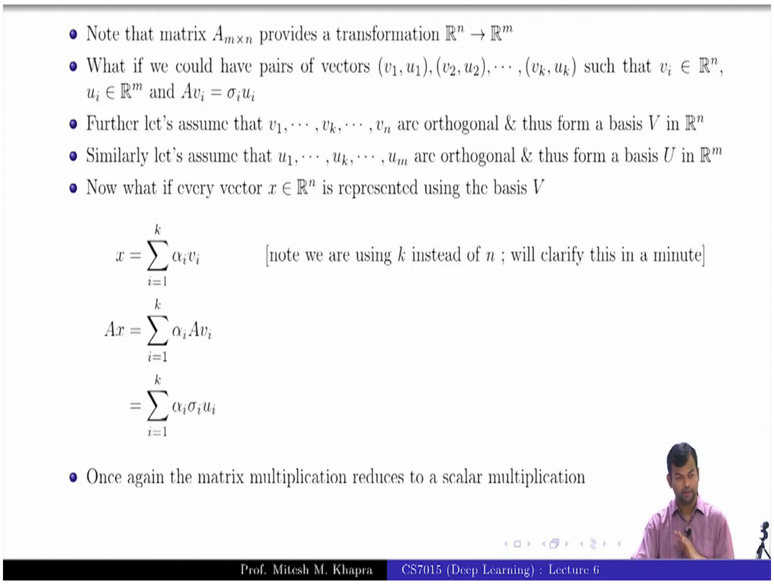
Student: (Refer Time: 06:01).

An m dimensional vector, right; hence they are super villain, right. Because they take the vector from one space and transform, it to a completely different space that completely lots lost it is identity, right. So, that is why rectangular matrices are even harder.

So now we just saw that for square matrices this eigenvectors form a very convenient basis where these operations reduce to a scalar operation. But now rectangular matrices do not even have eigenvectors. So, then cannot we have the same advantage there? Can we have the same advantage there? You can not, right because you do not have an eigenvector, but I would teach you about singular value decomposition. So, I better have something; so, get the connection, ok. There is a problem with square matrices with the rectangular matrices, ok. So now, let us see. So, we will try the aim is to see if we have something equivalent to this scalar transformation that we had for square matrices, ok.

How many of you have seen this in linear algebra before? So, you know whatever I am going to talk about, fine. So, the result of Ax is a vector belonging to R m and the original x belongs to R m. So, we do miss it miss out on this advantage that you could have reduced the matrix operation to a scalar operation, and now we will try to see if we can still get back that advantage, ok.

(Refer Slide Time: 07:32)



So, notice this is matrix, you can think of it as a function; which provides a transformation from Rn to Rm ok. So, what is the set of inputs to the matrix? It is vectors belonging to Rn, right that is the set of input.

Now, suppose we had a pair of vectors v 1 u 1 v 2 u 2 vk u k, each belonging to these 2 different universes one is Rn the other is Rm, ok. And there was a specular relation between them that A into vi is equal to sigma into ui, suppose I am just being ambitious let us see whether we can actually have this pair, but suppose we had this pair, then can you connect this back to the discussion on scalar operations, ok. So, let us just see that in detail, and we will of course, assume that these are orthogonal and form a basis. So, the vi is form a basis in Rn and the uy is form a basis in Rm, ok. Is that clear that is all straightforward, we have these vectors.

Now, every a vector belonging to Rn which was the input space can be represented using a linear combination of v straightforward, and any vector belonging to the output space can be represented of.

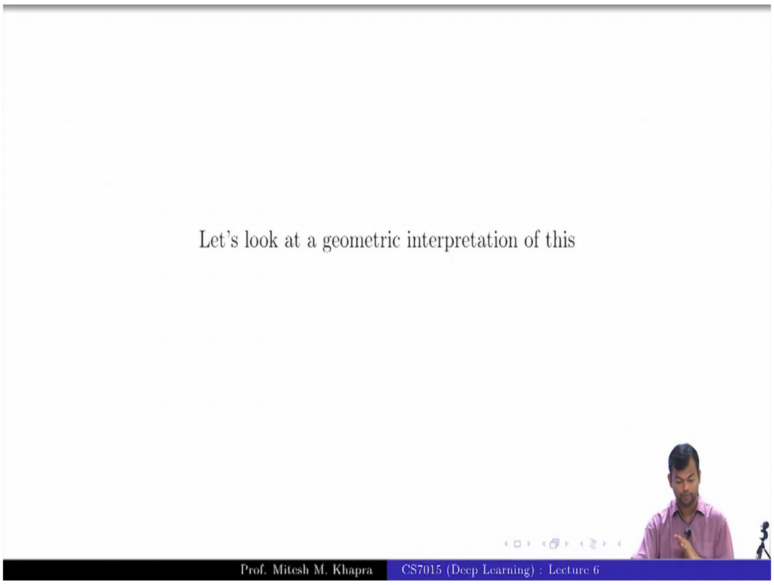
Student: (Refer Time: 08:55).

Of u, right, so, that means, any x in the input space I can write it as this linear combination. And now if I do the matrix operation what happens?

Student: (Refer Time: 09:09).

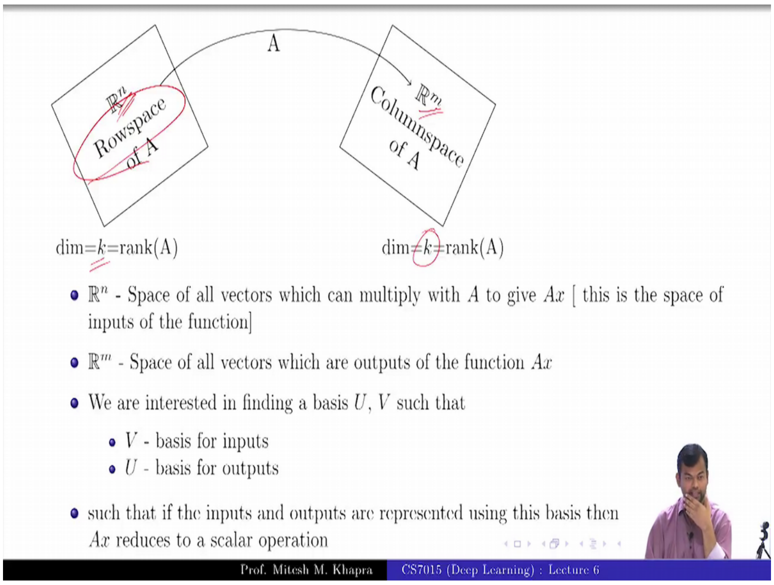
You get this A into vi, what is A into vi? Sigma ui sigma ui, I have still not shown you how to find these sigma is ui by the way, right, ok? Once again the matrix multiplication reduces to a scalar multiplication, ok.

(Refer Slide Time: 09:30)



So now let us try to look at a geometric interpretation of this.

(Refer Slide Time: 09:32)



So, what you have is this original space which is Rn you are using a as a matrix operation, right as a function and you are transforming vectors from n to Rm right. So, this is the space transfer that I was saying it vectors are being picked up from Rn and being put into R m, ok. And Rn is a space of all vectors which can act as inputs to this function, and Rm is a space of all vectors which are the outputs of this function Ax, ok.

Now, we are interested in finding a basis u v such that v is the basis for the inputs, when I say basis all of you should immediately start thinking of dash vectors.

Student: Orthonormal vectors.

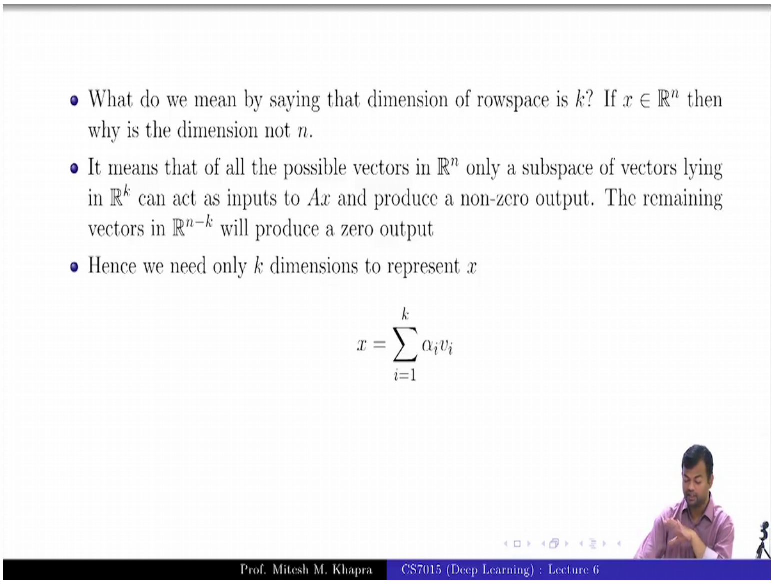
Orthonormal vectors orthogonal or orthonormal, right; Once we have orthogonal we do not care about the rest, u is the basis for the outputs; such that if the inputs are and outputs are represented using this basis, then all our matrix operations reduce to scalar operations. So, we are just trying to find the rectangular analogy for the square a phenomenon that we observed, ok. That is what were trying to do, now can you tell me, I have told you that if such a vn u exists, then you could do this. Can you give me such a un?

So, what do we mean by so, here I said actually I said this, right, that the dimension of the row space is actually k and the dimension of the column space is also k what do you mean by the dimension is I mean, right. Here I am telling this is Rn and this is Rm, and now I am telling you the dimension is k, what do I mean by that?

Student: (Refer Time: 11:05).

The only k linearly independent vectors, fine.

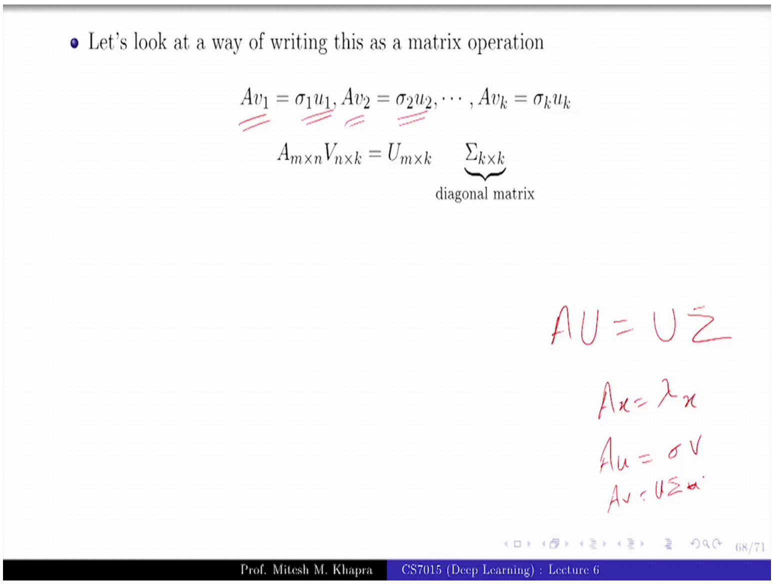
(Refer Slide Time: 11:11)



And this is again something from linear algebra which I expect you to know is that all possible vectors in Rn only a subspace belonging to Rk can actually act as input to a x to produce a non-zero output. So, I am talking about a null space column space and things like that right. So, this should be clear if it is not it is, ok. It is not very important at for us right now right.

And hence we have only k dimensions, fine.

(Refer Slide Time: 11:38)



So, let us look at a different way of writing this. So, you have this a v 1 is equal to sigma 1 u 1 Av 2 is equal to sigma 2 u 2. So, I can again do the same trick that I put all the vs into one matrix; where vi is are the columns of this matrix. And I will put all the us into another matrix where uis are the column of this matrix, is that fine everyone, ok? So far and then I can write it as this matrix operation. Same thing that we did when we are doing eigenvalue decomposition, right; So, we had written it as A into u is equal to u into sigma, right. Because there we had the condition that Ax is equal to lambda x, now we have a u is equal to sigma v or rather the other way around. So, Av is equal to did I missed up, did I no, right?

Student: (Refer Time: 12:29).

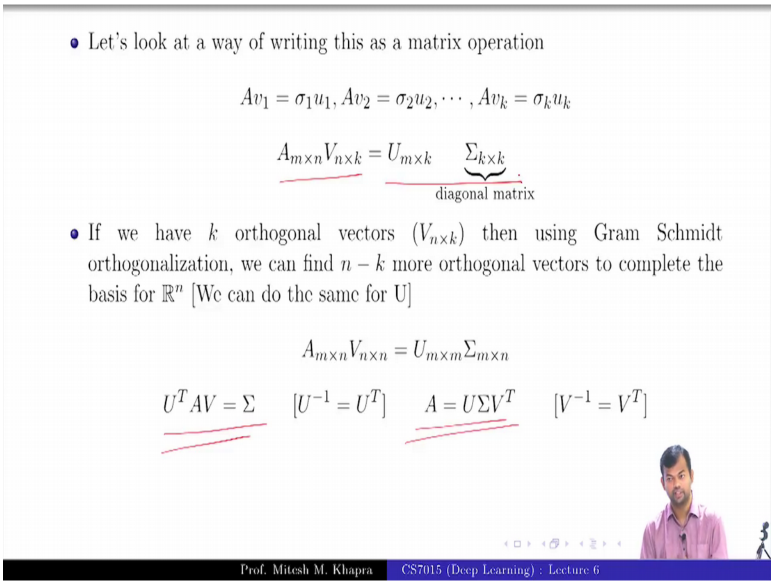
Sorry.

Student: (Refer Time: 12:32).

Fine, yeah so, Av is equal to sigma into u, ok.

So, is this fine no no, but when you do the diagonal operation you will get it as u into sigma y. The same way as a x equal to lambda x, but when you write it is A into u is equal to lambda comes later on right. So, everyone is fine, right? Can I go ahead, ok?

(Refer Slide Time: 12:56)



And if we have k orthogonal vectors; so, remember I said that this basis consists only of k dimensions, right. Because that is R the set of vectors which can act as input to A. So, what I, but I want a basis for the full Rn. So, what do I do for the remaining n minus k? Have you heard this gram schmidt orthogonalization, right. So, if I give you if there if you are trying to construct a basis for n, ok. For Rn rather and if I give you k orthogonal vectors, they can do k you can construct the remaining n minus k using Gram Schmidt orthogonalization right. So, you can get the full basis, fine. So, let me just see and this is orthogonal ok. So, you can write so, you see these 2 forms can you relate it to something that we have seen before in the course.

This is singular value decomposition, what else did we see before?

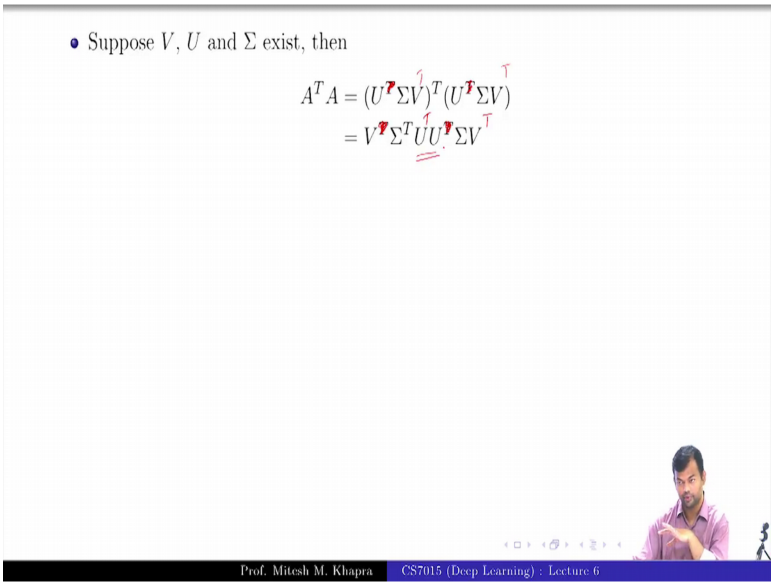
Student: (Refer Time: 13:53).

Eigenvalue, so, this exactly the same forms, right; And I have used the same set of tricks to arrive at it right. So, I first put the vectors into a column as columns into a matrix then wrote this in the matrix format, and then pre multiplied post multiplied by certain things and I got these 2 formats. And remember that v and u both are dash matrices.

Student: (Refer Time: 14:13).

Orthogonal matrices, right. So, that inverse is just their transpose, so, so far everything is fine, now I still do not know what U and V are, all this analysis is assuming that I know what U and V are. So now can you tell me how to get these Us and Vs.

(Refer Slide Time: 14:30)



Suppose v u and sigma exist, ok; then we can write this right. So, A is u sigma v so, A transpose would be the transpose of that. Now can you work with me, what is the next step?

Student: (Refer Time: 14:47).

Ok next.

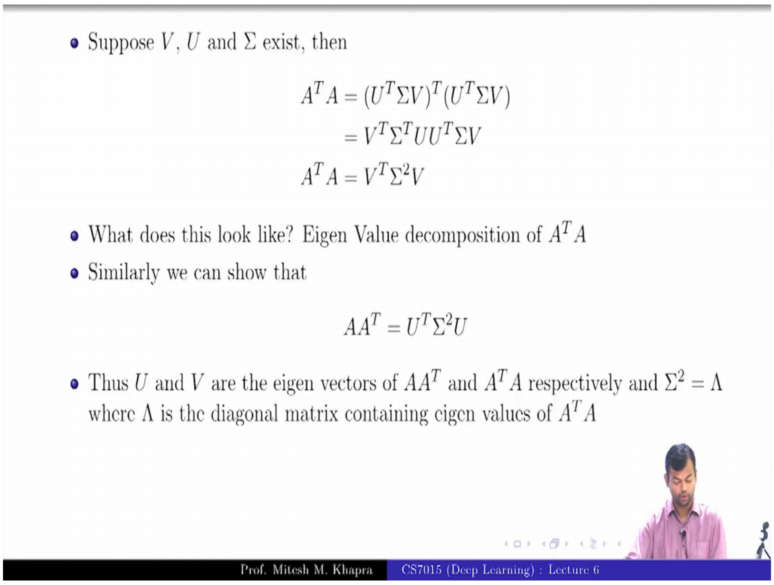
Student: (Refer Time: 14:51).

Oh ya, ok, this is u sigma v transpose. So, then this would be I think the next step is no the next step is also wrong, that fine? Ok fine, I just had some error with the transpose, ok. What will happen now? What will disappear from here?

Student: (Refer Time: 15:26).

U transpose U that is I write because U transpose the inverse of U.

(Refer Slide Time: 15:30)



So, you get this, what does this look like? This looks like the eigenvalue decomposition of.

Student: A transpose A.

A transpose A; that means, v consists of the.

Student: Eigenvectors.

Eigenvectors of the.

Student: A transpose A.

A transpose A; So now, can you tell me what u would?

Student: (Refer Time: 15:54).

Ok fine.

So, this looks like the eigenvalue of eigenvalue decomposition of A transpose A. Similarly, we can show that a A transpose is equal to u transpose sigma square u, ok. So, then u is the set of Eigen vectors of a A transpose, right. And now here what was with will the eigenvalue decomposition always exist for a matrix.

Student: No.

No under what conditions would it exist? First of all it has to be a square matrix.

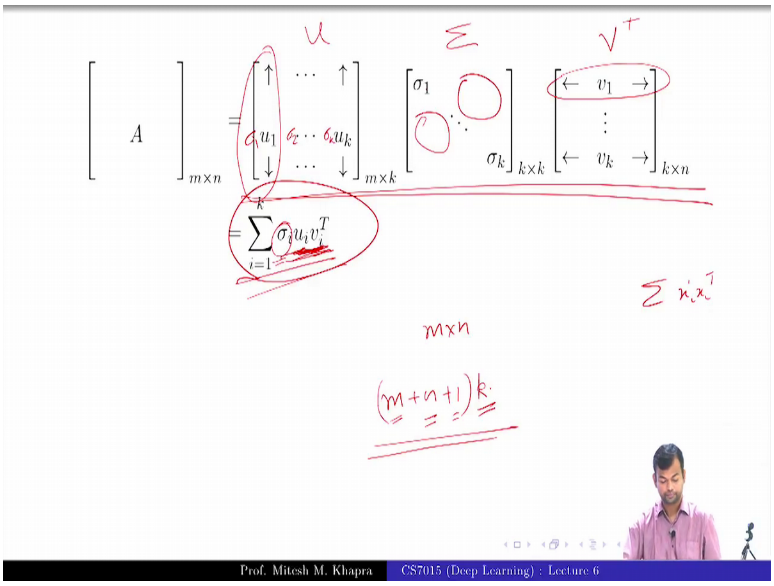
Student: (Refer Time: 16:32).

Ok right, but now for a rectangular matrix would be singular value decomposition always exist.

Student: Yes.

Yes, right. Because it depends on the eigenvalue decomposition of square symmetric matrices, ok. Is that fine, ok? So, for any matrix shall always have the Eigen value oh sorry the singular value decomposition, ok.

(Refer Slide Time: 16:58)



So, this is perhaps ya, ok, now just one last bit and let us see if all of you can understand this. So now, I can write A in this form this is nothing but what I already said, right. This is u this is sigma this is v transpose, ok. Now from here from this step do you see how I got to this step? This is something that we were struggling with yesterday also, when we were trying to find out summation x i x i transpose something similar here, you know the 4 ways of multiplying matrices, right. So, this is which way one of the ways, ok.

Ya so, does everyone get this, right. Ok. So, a simple thing would be first to just take these sigmas inside, right. Because this is a diagonal matrix, right this is all 0’s. So, these are actually you can write it as sigma 1 u 1 sigma 2 u 2 and sigma k u k, right.

Now, this ends up being the product of 2 matrices, right. And you can write it as a sum of columns into rows right. So, what I am writing it as a sum of sigma 1 u 1 multiplied by v 1. So, sigma 1 u 1 into v 1 transpose is a scalar matrix vector matrix right. So, each of these terms here is a.

Student: Matrix.

Matrix, and you are adding k such matrices, ok. Now try to relate it to reconstruction error. You are taking a matrix trying to write it as sum of many matrices. If I trim some terms from this some terms from this sum what would happen? If I have all the terms, then what would happen? I will get a back exactly. If I drop some terms what would happen?

Student: (Refer Time: 18:57) I get an approximation of A, how good would that approximation be?

Student: (Refer Time: 19:03).

First is depending on the number of dimensions, but is there a natural ordering in these dimensions if I want to throw away some dimensions which one would I throw away.

student: (Refer Time: 19:11).

Smallest.

Student: Singular values.

Singular values, sigmas are the singular values. So, you see that this is getting multiplied here, every matrix is getting multiplied by the singular value corresponding singular value. So, if I drop out the terms which have the smallest singular values, then those matrices the elements would have been very small. So, I will not lose much in the approximation right; so, again the same idea that I am trying to approximate the original matrix by a smaller rank, right.

By of so now, the original matrix had m cross n entries, ok. How much if I use only k eigenvectors or the sorry k singular vectors or k dimensions to approximate it, how much storage would I need? How many values do I need? So, the original matrix was m cross n. How many entries are there here?

Student: (Refer Time: 20:06).

Each of this is how much?

Student: (Refer Time: 20:10).

M for ui plus n for vi, ok. And plus 1 for the sigma, and how many of these are there?

Student: k.

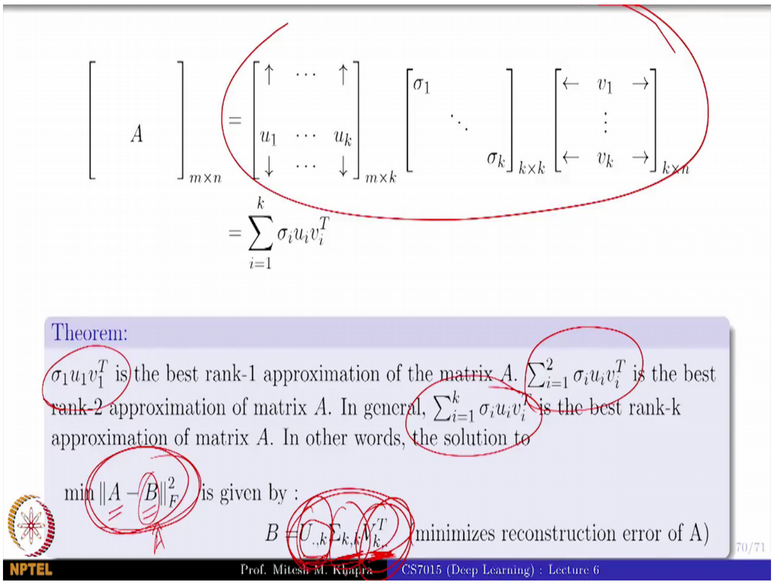
K so, if k is very less than your m and n, right. Then again you will have some compression you get this ok. So, all of these ideas are related and I want you to be able to connect them, right. That all of this is towards doing some approximations reconstructing some reconstructing a matrix from it is components, and doing this reconstruction in a manner that you end up making minimum error in the reconstruction. Is this idea clear? Even if some part of the math is not clear, is this idea clear? How many forget this? Ok so, some of you do not, you do not?

Student: (Refer Time: 20:59).

Ya so, what is the original dimension of A? M cross n, right now I am trying to reconstruct it using a sum of sum k terms, ok. So, hence this k comes, now each of these terms how many elements do I have? I have ui which is of dimension m, I have v i which is of dimension n and then I have the sigma I which is of dimension one, right. And I have k of these. So, this is the total amount of storage that I need. I am saying that as k is much less than m and n which would typically be the case, right.

Then you are getting a much lower space reconstruction of the original data, right. And you are doing this reconstruction smartly, because you are not taking any k dimensions, you are taking the k most important dimensions, and this most important is defined by the singular values, this is designed by the sigma is that fine?

(Refer Slide Time: 22:02)



Ok and actually there is a formal theorem which says that sigma 1 u 1 v 1 transpose is the best ranked one approximation of the matrix is; this a rank one matrix A sigma 1 u n I hope you guys have done the assignment, right. Sigma 1 u 1 v 1 transpose is the rank one matrix. And if I take this idea further, this summation is the best ranked 2 approximation and if I keep going, this summation is the best rank k approximation. So, what it says is that if you are trying to reconstruct the original matrix, right from these components. And if you go by the Eigen or the singular values, and you pick the ones corresponding the top k singular values then the best that is the best possible reconstruction that you could have done, ok.

Now, how do you formally define reconstruction? How would you make it as an optimization problem? What are you trying to minimize?

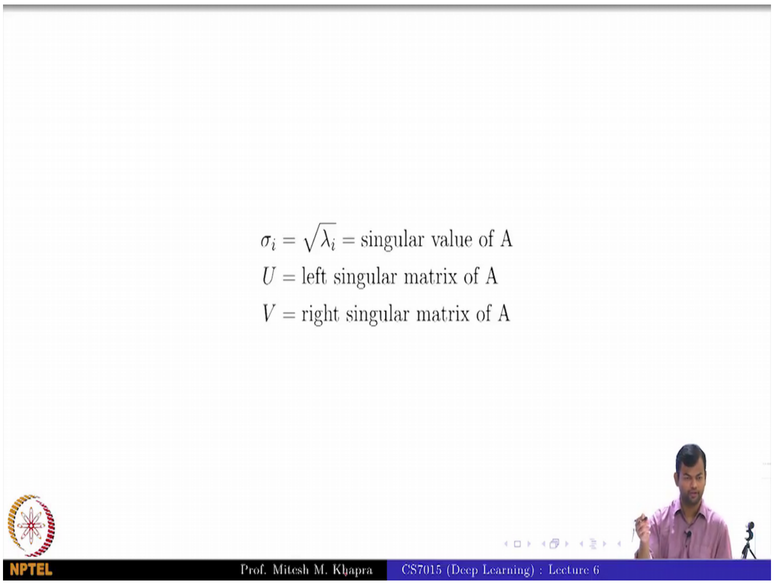
Student: (Refer Time: 22:59).

The actual matrix has some values which is the matrix A, ok. B is the reconstructed matrix using only k dimensions, how many of you understand? What is this product saying what is this?

Student: (Refer Time: 23:20).

First k columns of u, ok; These are the k singular values, and these are the first k rows of. So, ok I was just talking about this is the first k columns of u, these are the k singular values are put across the diagonal, and this is the first k rows of V transpose, ok. And this is exactly the product which I showed you here, is that fine?

(Refer Slide Time: 23:49)



Ok so, there is a theorem this is called the svd theorem it says that, if you want to reconstruct a then this is the best rank k approximation that you can get. Now if I want to paused it as an optimization problem what will I say what would I have minimized actually. This is the reconstruction, right. So, let us call it a hat actually, and what does this mean? This is the dash norm.

Student: (Refer Time: 24:15)

Provenience norm, what does the provenience norm give you? Squared difference between the elements, right roughly speaking ok; so, it will tell you what is the square difference between the ijth element of a and the ijth element of B. So, whenever we have this situation if you are trying to if this is our objective function that we have trying to reconstruct A, or try to transform something and get a predicted A or A reconstructed A then the best possible reconstruction would be given by this solution.

So, this optimization problem has a solution that you just use the eigenvectors of xx transpose and sorry a A transpose And A transpose A, right. Is that clear? Ok so, this is the theorem that we will be using when we are talking about autoencoders, and we will try to connect auto encoders to pcm ok. So, just revise this is the prerequisite for next class whatever we have done in the last 3 sort of extra lectures, you have to revise it before you come for class tomorrow, right, ok. And yeah this is sigma is just some terminology sigma is actually the square root of lambda I that was obvious, and U is called the left singular matrix of a and V is called the right singular matrix of A.