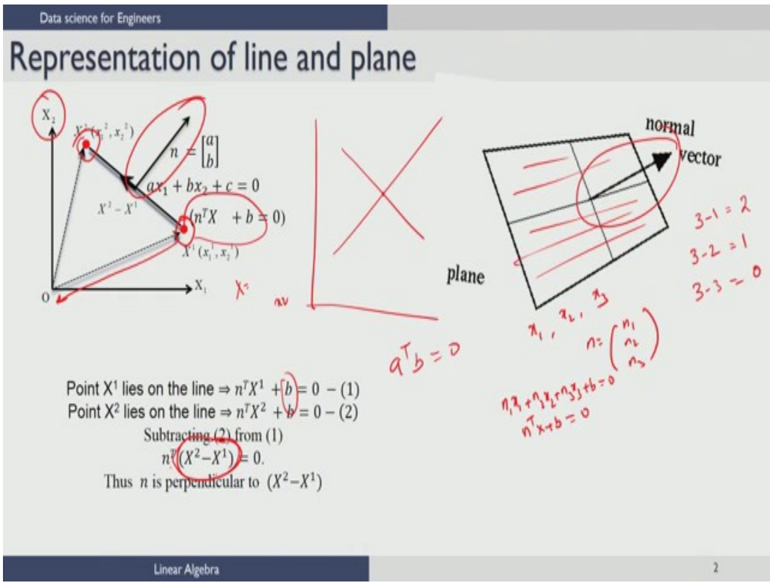
Data Science for Engineers
Prof. Raghunathan Rengasamy
Department of Computer Science and Engineering
Indian Institute of Technology, Madras

Lecture - 16
Linear Algebra - Distance, Hyperplanes and Halfspaces, Eigenvalues, Eigenvectors (Continued 1)

So, let us continue with our lectures on linear algebra for data science. We will continue discussing distances hyperplanes, half spaces, eigenvalues and eigenvectors in this lecture, and the lecture that follows this lecture. So, what we are going to do is we are going to think about the equations in multi-dimensional space. And then think about what geometric objects that these equations represent.

(Refer Slide Time: 00:49)



So, let us look at what equations mean from a geometric viewpoint. To do this, let us start with 2 dimensions. Let us assume that we have a 2 dimensional space in X_1 and X_2, and let us also assume that we have one equation that relates the variables; X_1 and X_2 which is $aX_1 + bx + c = 0$.

So, we want to understand what geometry this equation represents. It turns out in this case of the 2-dimensional space, this equation represents a line which is depicted here. So, a single equation in a 2-dimensional space represents a line. So, one might ask what does 2 equations represent.

And to understand this if you look at a picture like this, let us say I have one equation which is a line, let us draw the other equation which is also a line. Then if both of these equations have to be satisfied, then that has to be this intersecting point. So, 2 equations in 2 variables represent a point if these equations are solvable together.

Now, if you have no relationships between these variables, then we would say that we are representing all the points in the 2-dimensional space. And there is no relationship that constrains these points to either lie on a line or be a single point and so on. Now let us look at this equation, and then rewrite it in a form that is generally used. And that form is the following which is $n^T X + b = 0$, n is this column vector that is defined here. X is a vector of variables X_1 and X_2. And if you do a one to one comparison between this equation and this equation, you will see that b = c. So, a general equation can be written in this form $n^T X + b = 0$.

Now, we want to understand what this n depicts in this example. Now if you look at this picture here, we have shown n as a normal to this line. And let us see why that is true. To see that, let us first start by looking at 2 points on the line. So, let us start with this point X_1 and then X_2. Notice that both these points are on the line.

So, when I substitute X_1 into the equation for the line, it should satisfy it which is what is shown here $n^T X_1 + b = 0$ and when I substitute X_2 into the line equation that should also satisfy. So, $n^T X_2 + b = 0$. Now what you could do is you could subtract the first equation from the second equation. The b's will get cancelled. You will have $n^T X_2 - X_1 = 0$. Now let us interpret this equation. From vector addition you know that if I have X_1, $- X_1$ is in this direction.

So, when I do $X_2 - X_1$ I am adding X_2 1 $- X_1$, which is equivalent to starting from here and going here ; which is basically through vector addition in the direction of this line. So, what this equation basically tells us is that this is in the direction of the line.

And from our orthogonality lecture we saw before, we said if $A^T b = 0$, then a and b are orthogonal. Since n^T this quantity is 0, and this quantity is in the direction of the line n has to be perpendicular to the line. So, this is a very important idea that we will use in data science quite a bit, when we looked at linearly separable classes. And then classifying classes that are linearly separable.

Now, if you want to extend this, and then ask the question, if I have one equation in a 3-dimensional space, what does that represent. Now the form of the equation will be very similar. You will have something like this here, and n now would become supposing you have 3 variables X_1 X_2 X_3. N could be n_1 n_2 n_3.

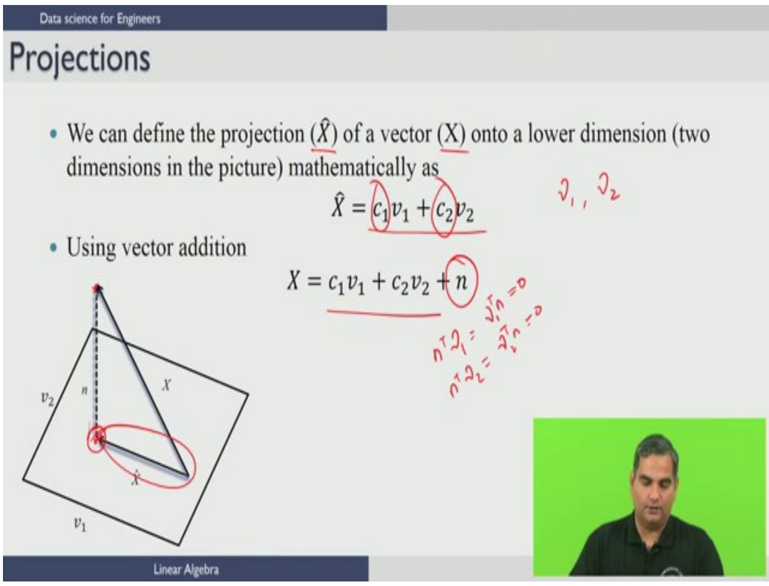
So, the same equation would be $n_1 X_1 + n_2 X_2 + n_3 X3 + b = 0$. So, which is $n^T X + b = 0$. So, irrespective of what-ever is the dimension of your system, you can always represent a single linear equation in this form $n^T X + b = 0$.

Now, we ask the question as to what a single equation would represent in a 3 - dimensional space. And in a 3-dimensional space a single equation would represent a plane a 2-dimensional object. The way to see this is in 3 dimensions we have 3 degrees of freedom. If you write one equation you are taking away one degree of freedom. So, we are left with 2 degrees of freedom.

And a 2 degree of freedom object is basically a plane. And this is what is shown here. So, if I have one equation, that equation itself would represent this plane right here; which is what we see. And very similar to how we drew the normal to the line here, this n would represent a normal to the plane. So, this would represent a projection outside out of the plane orthogonal to the plane. So, that is what n would represent.

So, in 3 dimensions, one equation would represent a plane. And in 3 dimensions 2 equations would represent a line, because we have 3 degrees of freedom if you take away 2, then you have one, a one-dimensional object is a line. And if you have 3 equations, then you have 3 - 3 = 0, the 0-dimensional object would be a point, and this would be a point as long as these 3 equations are consistent and solvable.

(Refer Slide Time: 08:10).



Now, that we have talked about what equations represent and so on. One of the things that we are quite interested in and you will see this again and again in data science, as we teach some of the algorithms later such as principle component analysis and so on. We are always interested in projecting vectors onto surfaces. The reason why we are interested in doing this is, because many times we might want to represent data through a smaller set of objects or a smaller number of vectors. So, in some sense the data cannot be completely represented by these vectors.

So, we might ask the question as to what is the best approximation for this data point based on the vectors that I want to represent this data point with. So, this is a very important question that we will keep asking again and again. You will understand this in much more detail and clarity, once we talk about some of the data science concepts.

For now, I am just going to treat this mathematically, and then explain to you how we do projections and what are the equations that we can get for writing down projections. The interpretation for this and the use for this in data science is something that we will see as we go along this course later. So, let us take a very, very simple example. Let us assume that I have a plane which is shown here in this picture.

Since I have a plane basically we are looking at A_3-dimensional space. A plane has to be represented by 2 dimensions, because it is a 2-dimensional object. Let us assume that the basis vectors for this plane are v_1 and v_2. We have already discussed what basis vectors are in a previous lecture. So, for a 2-dimensional object we will need 2 basis vectors.

So, let us assume these basis vectors are v_1 and v_2. Just to recap what this basis vectors are useful for is that, any line on this plane, basically can be written as a linear combination of v_1 and v_2. That is what we described before, that these basis vectors are enough to characterize every point or any vector on this 2-dimensional plane. So, any vector can be written as a linear combination of v_1 and a v_2.

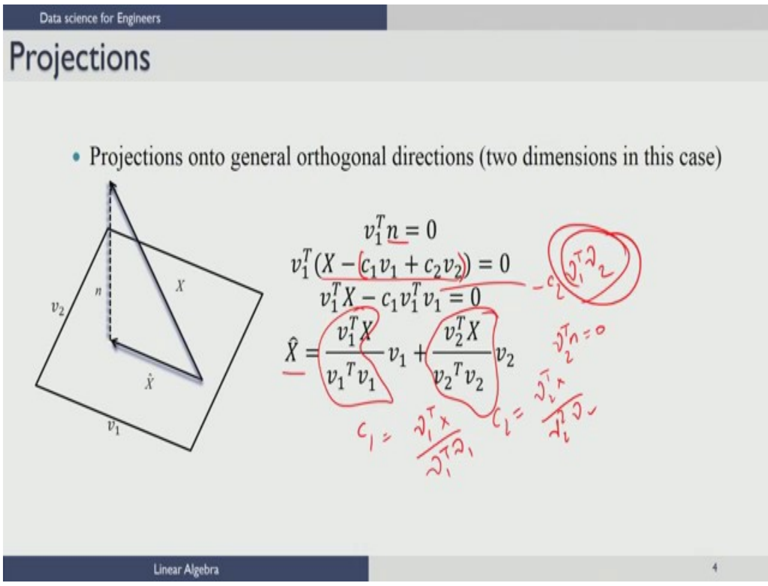
Now, the way this picture is drawn, you would see that this is the plane and I have let us say, a vector that is coming out of the plane. So, this is not clearly in the plane. So, it is projecting out. So, from the data science viewpoint if you want to make an analogy, what we are saying here is that, I have a data here X which is represented by this vector. I want to write this simply as only a function of v_1 and v_2. So, in other words I want to represent this vector X, in a tool, I cannot do it exactly projecting out of the plane. So, I might ask what is the next best thing that I could do in this case. It turns out the next best thing to do would be to project this vector onto the plane, because ultimately, however I write this vector with only this 2 basis vectors it has to be on the plane.

Now, there are many vectors on the plane. I want to find what is the best projection for this onto this plane. So, a common sense idea would be to say, I want a point here which I write. And if this is the projection of this vector I want this distance to be minimized. So, you can see why that is. Think about this if you keep projecting it bac_k to the plane, if this is the closest point if the vector is already in the plane, it would be the same vector that is also the product right. So, as soon as this vector goes up slightly outside the plane, I want it to be projected bac_k. So, that it is closest to that point of projection. So, how do we explain these concepts mathematically? So, we do that here. First,$\hat{X}$ is the projection of X onto lower dimension in this case 2 dimensions.

And since $\hat{X}$ has to be in the lower dimension, We already know that it can be written as a linear combination of v_1 and v_2. So ,$\hat{X}$ is c_1 v_1 + c_2 v_2. The c_1 and c_2 are yet to be determined. So, we do not know what those are. We are going to try and determine these 2 using this idea of projection. So, what we are going to say is that, if this is the projection, then the closest point from here would be when I draw a perpendicular or drop a perpendicular onto the plane. So, as long as these 2 points when I connect by vector n, that vector is perpendicular to this plane, then I would have found the closest point on this plane, which is what I am going for in terms of projections.

So, using vector addition, again we can start from here let us say, and this is x. So, X can be written as $\hat{X}$+ n, which is what is written here. And $\hat{X}$ has been expanded to be c_1 v_1 + c_2 v_2. Notice that while we write this, the fact that we are using a projection comes from this n being perpendicular to the plane. So, what does n being perpendicular to the plane mean? If n is perpendicular to the plane, then we know that n^Tv_1 or v_1^T n both are the same will be 0. Similarly, $n^Tv_2 = v_2^Tn$ will also be = 0. So, these are 2 facts that will know, if n is perpendicular to the plane. So, how are we going to use this to calculate c_1 and c_2 is what I am going to show you in the next slide.

(Refer Slide Time: 15:09)



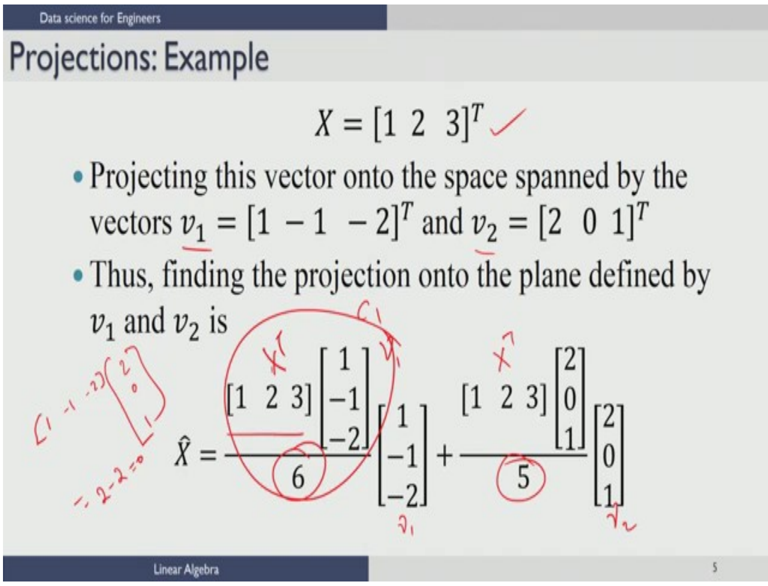
So, let us first take this $v_1^T n = 0$; the first equation I wrote. Let me write n as this from the previous slide, because X was c_1 $v_1 + c_2$ v_2 + n, I simply move c_1 v_1 and c_2 v_2 to the other side. And I have this equation right here.

Now, when I expand this equation, I will get $v_1^T X$ - c 1 $v_1^T v_1$ and I will also have another term which would be here - c_2 v_1^T v_2. Now as the first case I am going to show you how you do projections on 2 orthogonal directions. Now if these 2 directions are orthogonal the basis vectors themselves are orthogonal, then we know that this will be 0, that is the reason why this term drops out. And I have $v_1^T X$ - c_1 v_1^T v_1 = 0. Take this to the other side, and then bring $v_1^T v_1$ to the denominator, then you will get $c_1 = v_1^T X$ divided by $v_1^T v_1$.

Now, you could use the same idea, and then do the calculations for $v_2^T n = 0$. And when you do this, again you use this fact that $v_2^T v_1$ or $v_1^T v_2 = 0$ because these are orthogonal directions. And then you will end up with this equation for c_2, which will be $v_2^T X + v_2^T v_2$.
Once you get this, then you can bac_k out the projection and the projection is c_1 times v_1 + c 2 times v_2. So, this is how you project a vector on to 2 orthogonal directions, and this can be extended to 3 orthogonal directions 4 orthogonal directions and so on. Because all you will get if let us say it is 3 orthogonal directions then you will get $v_1^T X$ $v_1^T v_1$ for c_1, this is for c 2 and $v_3^T X$ divided by $v_3^T v_3$ for the third constant c_3.

So, this is how you do projection. This is a very, very important idea, and this will be used in many many places in data science. So, it is worthwhile to clearly understand this.

(Refer Slide Time: 17:52)



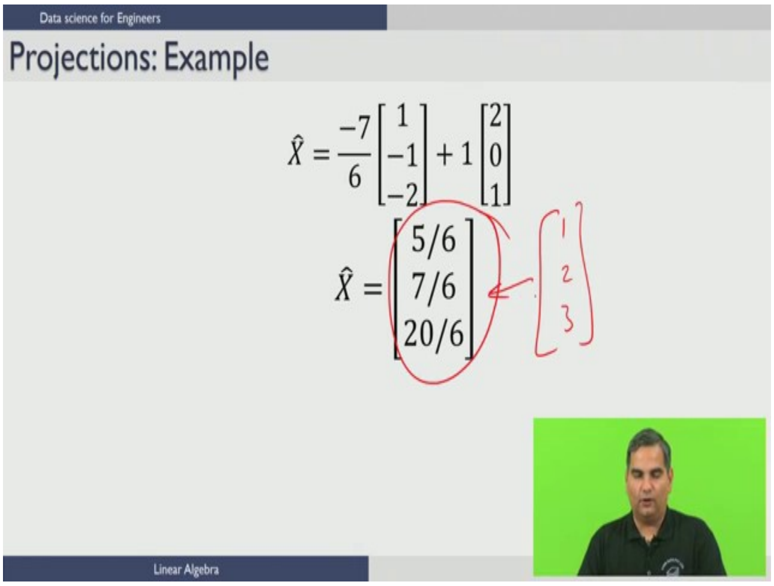
Now, let us move on to doing an example for this projection. Let us take a very simple example, let us say I have vector 1 2 3 transpose; so, column vector. So, this is a vector in a 3-dimensional space.

Now, let us take 2 vectors and then make a plane. So, let us take vector v_1 which is 1 - 1 - 2, and v_2 which is 2 0 1. And then try and see whether I can project X on to these. Let us first find out whether these 2 vectors are orthogonal. So, to do that we have to do $v_1^T v_2$. So, I am going to do 1 - 1 - 2 2 0 1. So, this will be one times 2 - 1 times 0 0 - 2 times 1 2. So, 2 - 2 is 0.

So, we know that these 2 vectors are orthogonal. So, we can use a formula that we had before. Now this formula is what we apply here. So, this is $v_1^T X$ transpose, sorry, this is $X^T v_1$ which is 1 - 1 - 2, and this should be $v_1^T v_1$. So, that will be one square + 1 square + 2 square. So, 1 + 1 + 4, 6. So, this is constant c_1 that we get, and this is multiplied by v_1.

And if you look at the second term here. So, this is X^T this is v_2 2 0 1. And this should be $v_2^T v_2$. So, this should be $2^2 + 0^2 + 1^2 = 5$. And we have this vector v_2.

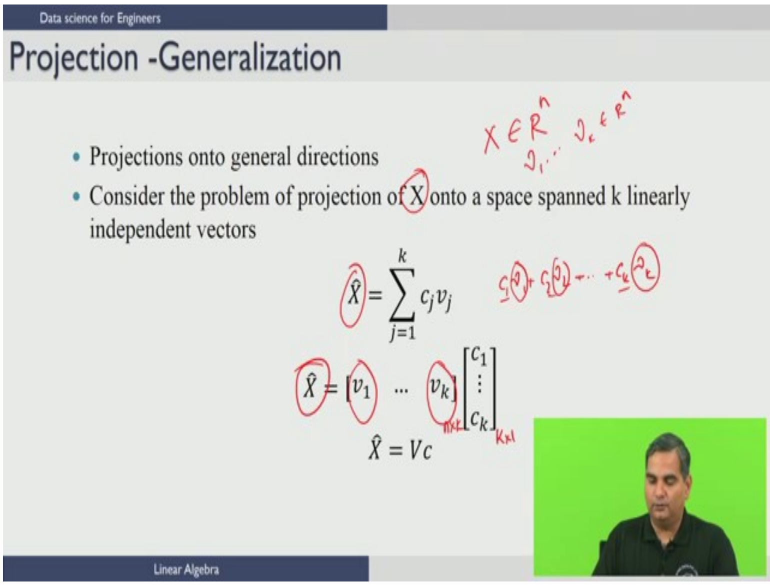
(Refer Slide Time: 19:50)



So, once you simplify this further you get the projection as the following. So, my original data vector 1 2 3, when it is projected onto a space spanned by these 2 basis vectors becomes this.

So, in other words, if I had a data point 1 2 3, and then I say I want to represent this with only 2 vectors that I had identified before whatever reason it might be, then the best representation is the following is what this projection says.

(Refer Slide Time: 20:26)



Now, we talked about projecting on to certain number of directions. And we also talked about projections when these directions are orthogonal. I am going to generalize this in the coming slides. So, that we have a result that is general and can be used in many places. So, I am going to look at how we can project this vectors onto general directions. So, let us consider the problem of projection of X onto space spanned by k linearly independent vectors,. Now I have dropped this notion of orthogonal here, I am simply saying these vectors are independent. As before since I want to project X on to k linearly independent vectors.

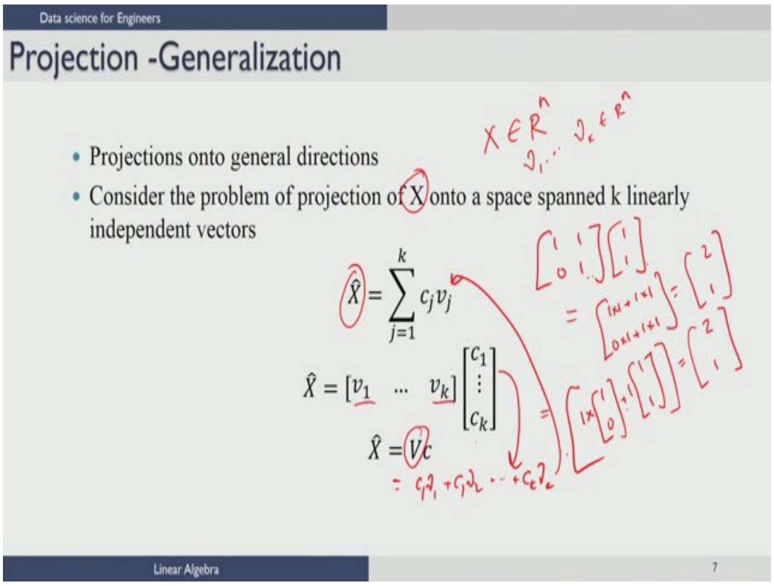
I am going to represent that projection as $\hat{X}$. And because this $\hat{X}$ is in a space spanned by this k linearly independent vectors, I can write this as a linear combination of these k vectors; which is what I have written here. So, if you expand this you will get c_1 v_1 + c_2 v_2 and so on + c_k v_k. Notice in this equation it is important to really understand this carefully. Notice in this equation v_1 v_2 v_k are all vectors, and c_1 c_2 c_k are scalar constants.

We can write this equation also in this form. Where, what we do is we $stac_k$ these vectors, into a matrix. So for example, if X is in an n dimensional space Rn. Then we would assume that each of these vectors are also in R n. So, v_1 to v_k are all element of Rn. And when you $stac_k$ k vectors like this in a matrix, then you would get a matrix of dimension n by k.

And since there are k constants, which I have put in a vector. So, this would be a vector of dimension k by 1. And you can notice that this n by k times k by 1 will give you an n by 1 vector which is what this $\hat{X}$ nonetheless, this n by one vector is in a space spanned by these k vectors linearly independent vectors.

Now this is an important thing to notice, if you go bac_k and then say let me expand this, then basically you should get this. And this is another way of thinking about matrix multiplications, which is important to understand. So, let me illustrate this with some very, very simple examples so that we use this at later times.

(Refer Slide Time: 23:25)

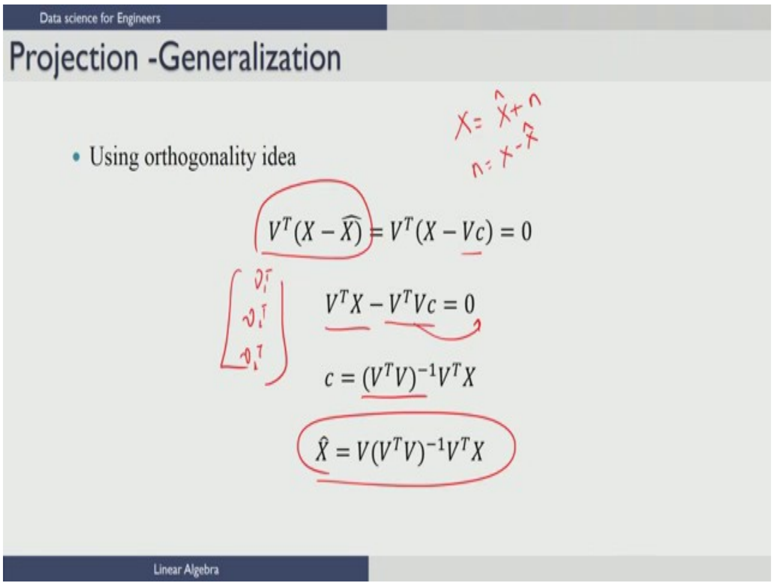


So, if I have let us say a matrix 0 0 0 1, I will say, and I want to multiply this by 0 0. The standard way of doing this would be one times one + 1 times one and 0 times 1 + 1 times 1.

So, this will be 2 1, right. This is your standard matrix multiplication that you have seen. You can also interpret this slightly differently. You could say that this matrix multiplication is also one times this vector 1 0 + 1 times this vector. So, this is what we can think of this, matrix multiplication as. Now if
you notice this will also give you the result 2 times 2 0 0. So, what we are doing is there are many columns here and there are, these scalar constants much like this. So, when we multiply this this will be c 1 times the first column; much like, how we have written here. So, this will be c 1 times v_1 + c_2 times v_2, all the way up to c_k times v_k.

So, this and this are same and you see that this is this. So, $\hat{X}$ can be written as v times c ; where v is a matrix where are all these basis vectors are stac$_k$ed in columns. And c are the scalar constants which have been stac$_k$ed as a single column. Now, let us proceed to identify the projection from here.

(Refer Slide Time: 25:00)



We then use the orthogonality idea. Remember, we have X =$\hat{X}$+ n. That means, n = X -$\hat{X}$ and if $\hat{X}$ has to be a projection, then n has to be a vector that is orthogonal to the space spanned by the k linearly independent vectors. For n to be orthogonal to a play to a geometric object spanned by this k linearly independent vectors. N has to be orthogonal to every one of these vectors.

So, that is what we write here in a matrix form instead of writing v 1 v_1^TX -$\hat{X}$ is 0 v_2^TX -$\hat{X}$ is 0 and so on. So, we write this in a matrix form where we say v^Tthere I will have $v_1^T v_2$ transpose. All the way up to v k^Ttimes X -$\hat{X}$= 0.

Now, $\hat{X}$ from the previous slide was v times c. So, we^{T}X - v c = 0. So, if I expand this I will get v^T X - v^Tv c = 0. If I take this term to the other side, and then do the inverse. I will get c = v^Tv inverse v^TX. Whenever we take inverses we have to always make sure that we can actually identify an inverse. In this case I will be guaranteed to have an inverse for v^Tv if the columns of v are linearly independent. And the fact that we have chosen these basis vectors as linearly independent already, assures us that those are linearly independent.

So, this inverse is something that exists. Once we calculate this c we know $\hat{X}$ is v times c. So, I simply plug this v c bac_k in and I get the expression for projection. So, this is how you do projection onto general directions. Now this is a very important idea that is used in several data science algorithms. In fact, this is a bac_kbone for something called principal component analysis. And this is also used in many many other machine learning algorithms.

So, it is important to understand this idea very clearly. Now that we have understood projections, in the next lecture I will describe the notion of an hyper plane and half spaces. And then continue on to eigenvalues and eigenvectors. I will see you the next lecture.

Thank you.