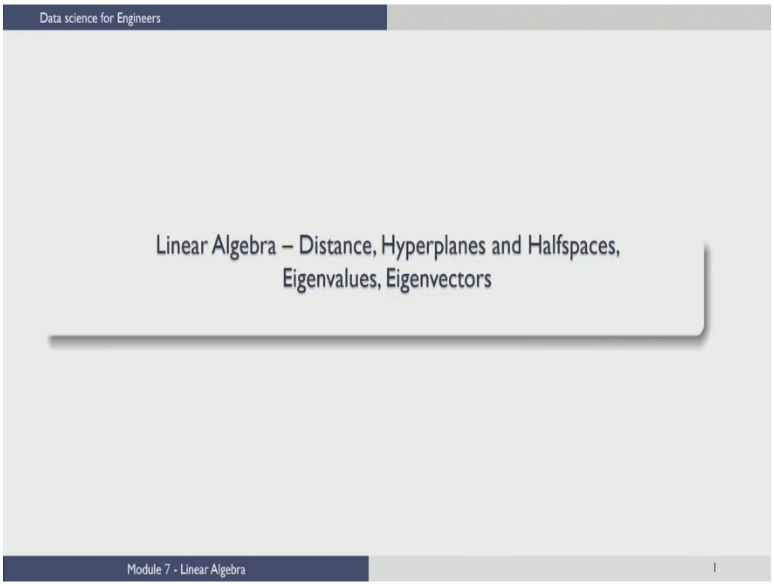
Data Science for Engineers
Prof. Raghunathan Rengasamy
Department of Computer Science and Engineering
Indian Institute of Technology, Madras

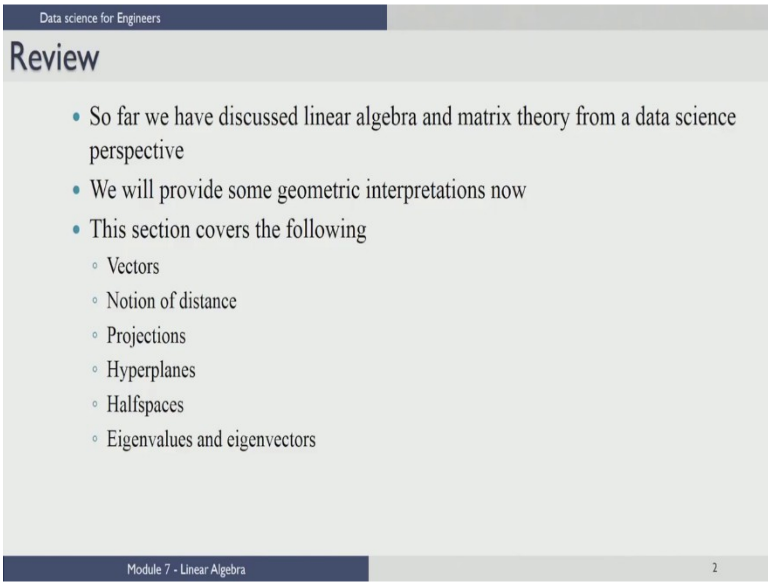
Lecture - 15
Linear Algebra - Distance, Hyperplanes and Halfspaces, Eigenvalues, Eigenvectors

(Refer Slide Time: 00:12)



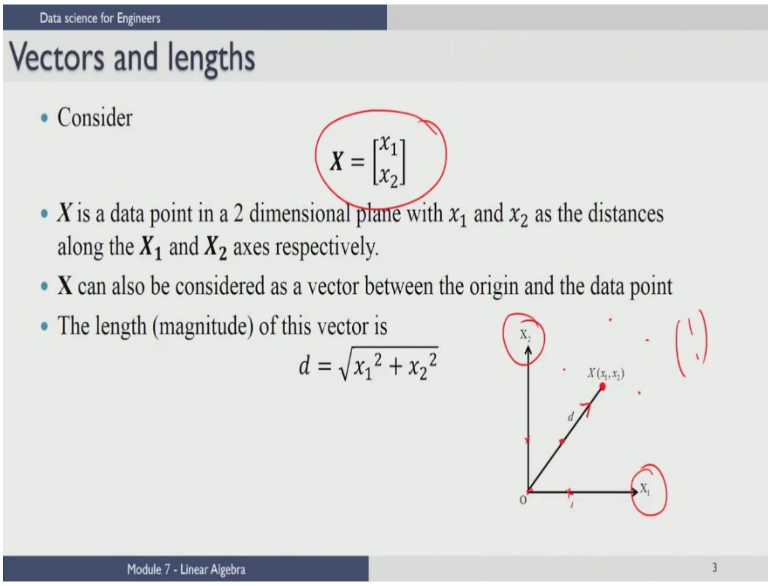
In the previous lectures we looked at linear algebra. But we took a linear algebraic view where we looked at equations and variables and solvability of these equations and so on. The same subject we could also take a geometric view, where we think about vectors and hyperplanes, half spaces and so on. So, we are going to cover that in the next couple of lectures that we are going to have on linear algebra. While we do this, we are going to cover the ideas of distance, hyperplanes, half spaces, Eigenvalues Eigenvectors. Now, some of these are things that would be very well known to most of you nonetheless, for the sake of completeness, I will go through all of these ideas and then I will use all of those ideas, when we describe hyper planes, half spaces and so on.

(Refer Slide Time: 01:21)



So, we will cover vectors notion of distance, we will talk about projections, we will talk about hyper planes, we will talk about half spaces and then we will talk about Eigenvalues and eigenvectors in this lecture. Till now, if we have been looking at a X = b and X as set of variables that needs to be calculated. So, we have been using this notation x_1 x_2 as a vector, where we have been interpreting this as a solution to a variable x_1 and a solution to variable x_2 and so on.

(Refer Slide Time: 01:37)

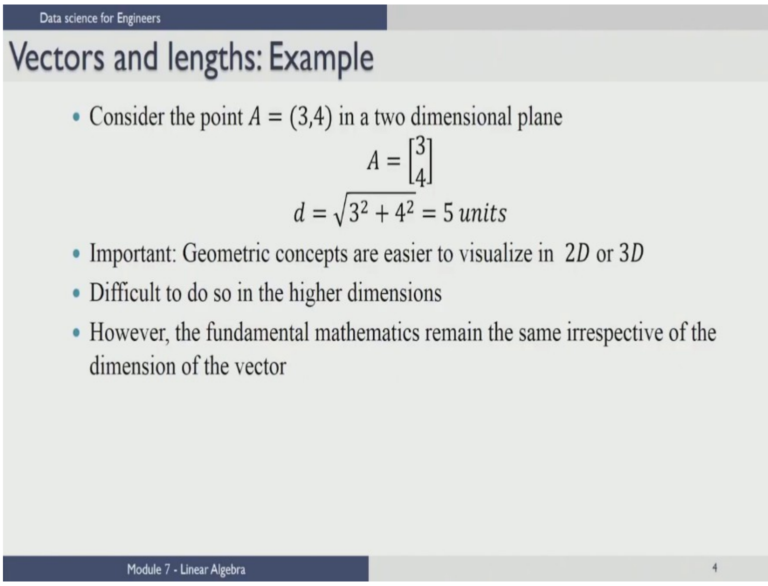


Another way to think about the same vector X is to think of this as actually a point in a 2-dimensional space and here we say, it is a 2 dimensional space because there are 2 variables. So, for example, if you take x_1 and x_2 you could think of this, as being a point in a 2 dimensional space, where there is one axis that represents x_1 and there is another axis that represents x_2, and depending on the value of the an x_1 and x_2 you will have a point anywhere in this plane.

So, for example, if you have let us say 1 as your vector, and if this is one and this is one, then the point will be here and so on. So, what we are doing here is, we are looking at vectors as points in a particular dimensional space. Since, there are 2 numbers here we talked about 2-dimensional space if for example, there are 3 numbers here, then it would be a point in a 3 dimensional space, you could also think of this as a vector and we defi ne the vector from the origin.

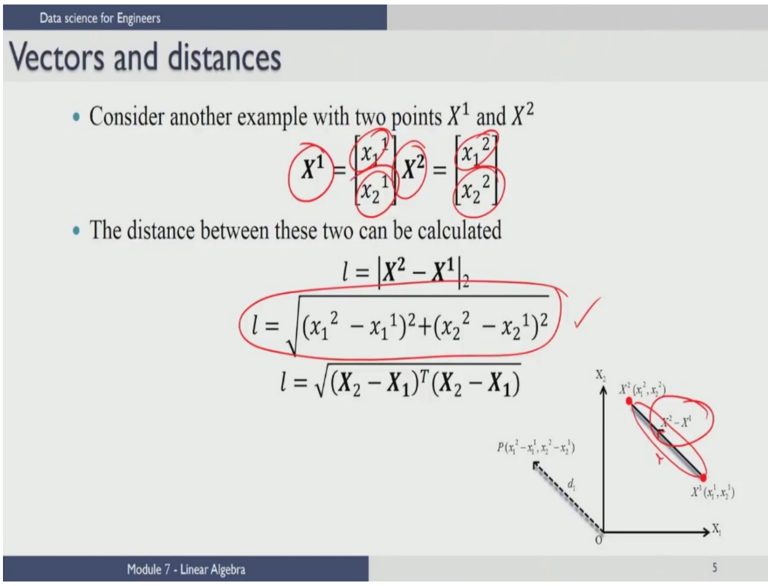
So, I could think of this X as a vector, where I connect origin to the point. So, this is another view of the same vector X and once we think of this as a vector then, vector has both direction and magnitude. So, in this case the direction is this and the magnitude is, what we think of as a distance from the origin and in this case we all know, this well-known formula for Euclidean distance, which is root of $(x_1^2 + x_2^2)$ right? So, that is the distance of this point from the origin.

(Refer Slide Time: 03:59)



Now, just as a very, very simple example, if you have a 0.34 then you can find the distance from the origin is root of $(3^2 + 4^2)$ is going to be = 5. It is important to notice that the geometric concepts are easier to visualize in 2D or 3D; however, they are difficult to do. So, in higher dimensions, nonetheless since the fundamental mathematics remain the same what we can do is, we can understand these basic concepts using 2D and 3D geometry and then simply scale the number of dimensions, and then most of the things that we understand and learn will be the same at higher dimensions also.

(Refer Slide Time: 04:45)

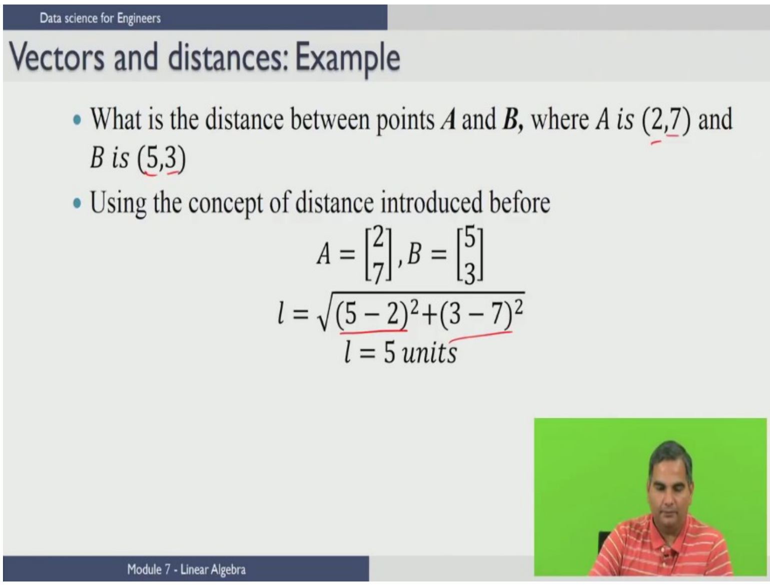


So, in the previous slide we saw one point in 2 dimensions. Now, let us consider a case where we have 2 points in 2 dimensions. We have x_1 here, which has 2 numbers representing the 2 coordinates and we have x_2 here, which also represents the 2 coordinates. Now, we ask the question as to, whether we can de ne a vector which goes from x_1 to x_2. So, pictorially this is the way in which, we are going to de ne this vector. What we do is, we draw a line starting from x_1 to x_2 and this vector is $x_2 - x_1$, the direction of the vector is given by this here, much like the previous case every vector will have a direction and a magnitude.

So, we might ask what is the magnitude of this vector and that is given by the wellknown formula that we see right here. Where what you do basically is, you take the x_1 coordinate of this point and this point take the difference square it, take the x_2 coordinate of this point the x_2 coordinate of this point, take the difference and square it add both of them and take a root and that is the equation that we have here.

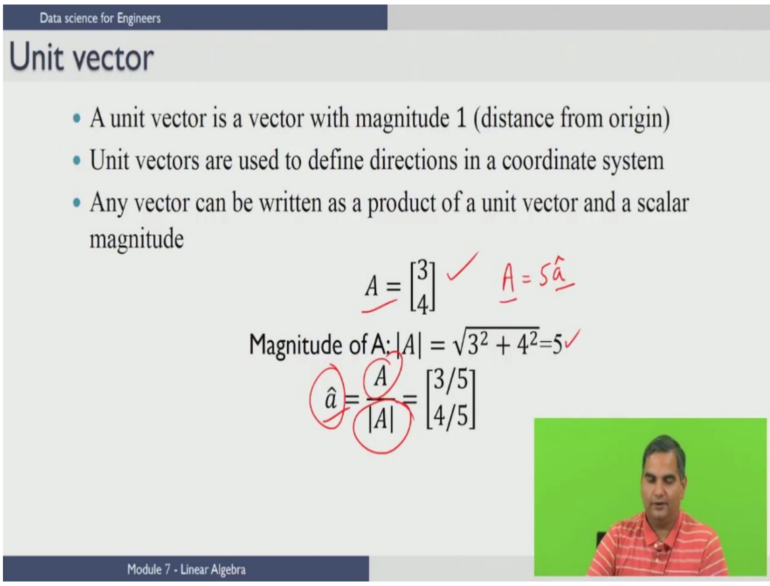
This is the length of this vector right here, this also can be written in a compact form as given here, which is root of $(x_2 - x_1)^T (x_2 - x_1)$.

(Refer Slide Time: 06:30)



Two simple examples to illustrate this, if I have 2 points A and B where A is 2 7 b is 5 3 then, the distances you take the difference between 5 and 2 and then square it and then, take the difference between 3 and 7 and then square it and then you will get your length as 5. So, that would be the length of the line that is, drawn between the 2 points A and B.

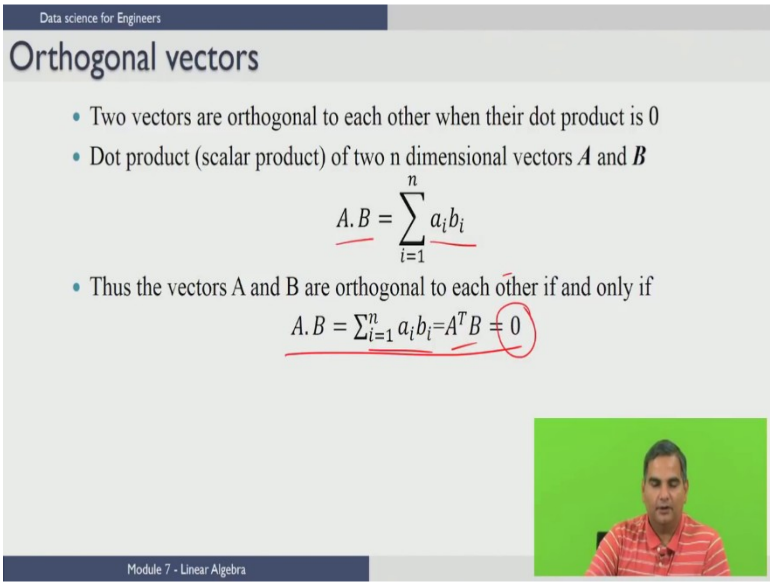
(Refer Slide Time: 06:59)



Now, it is useful to de ne vectors with unit length, because once you write a vector in unit length any other vector in that direction, can be simply written as the unit vector times the magnitude of the vector that you are interested in.

So, how do I define a unit vector, let us take this vector a 3 4, we know that the distance from the origin for this vector is root of $(3^2 + 4^2) = 5$. So, to define a unit vector what you do is, you take the vector and divide it by the magnitude of the vector. So, in this case it is 5. So, the unit vector becomes 3 by 5 4 by 5. So, the interesting thing is that, this unit vector is in the same direction as a; however, it has magnitude 1. So, I could write A itself as 5 times A. So now what has happened is this is a unit vector and this a has magnitude 5, which is what we derived here.

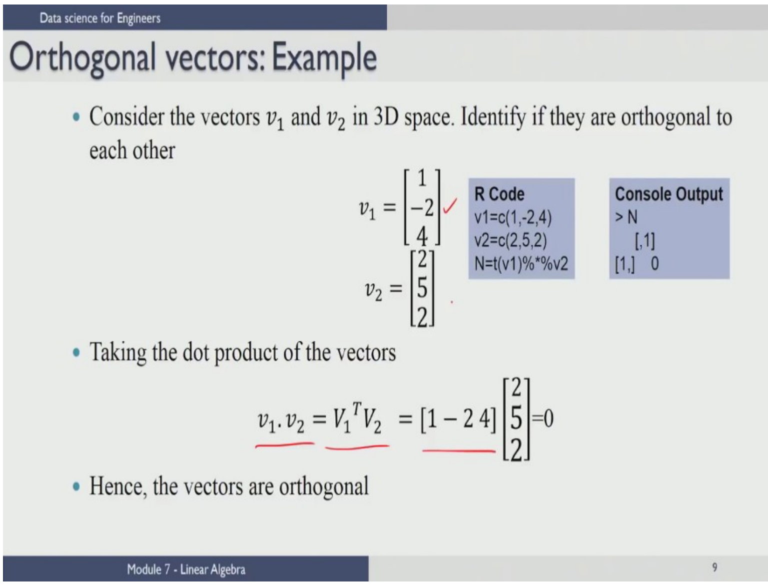
(Refer Slide Time: 08:04)



We introduce the next concept, which is important for us to understand many of the things that we are going to teach. If there are 2 vectors, we call these vectors as orthogonal to each other when their dot product is 0. So, how do I de ne the dot product? So, if I take 2 vectors A and B A dot B is simply $\sum_{i=1}^{n} ai\, bi$.

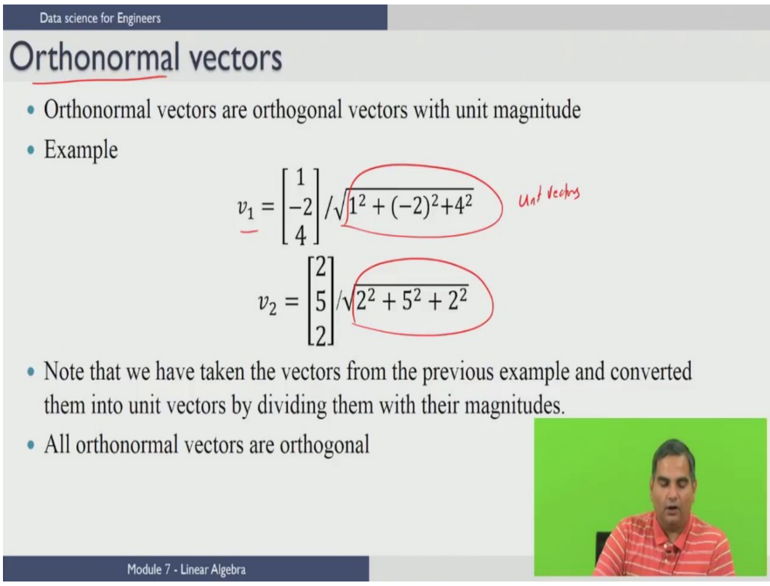
So, basically what you do is, if you have 2-dimensional vector then you take the 2 x coordinates multiply them, and then you take the 2 y coordinates and multiply them and add both of them you will get the dot product. This dot product again much like the distance that we saw before, can also be written in a compact form as A^T B you can quite easily see that this and this will be the same, and if this dot product turns out to be 0 then we call this vectors A and B as being orthogonal to each other.

(Refer Slide Time: 09:06)



So, let us take an example to understand this, let us take 2 vectors in 3-dimensional space. Let us say, I have one vector which is 1 - 2 4 and I have the other vector which is 2 5 2 and if I take a dot product between these 2, which is $v_1^T v_2$ or $v_2^T v_1$, both will be the same. I have v_1^T which is 1 - 2 4 and this is 2 5 2, if this will be one times 2 - 5 times 2 + 4 times 2 you will see that goes to 0. So, we say that these 2 vectors are orthogonal to each other.

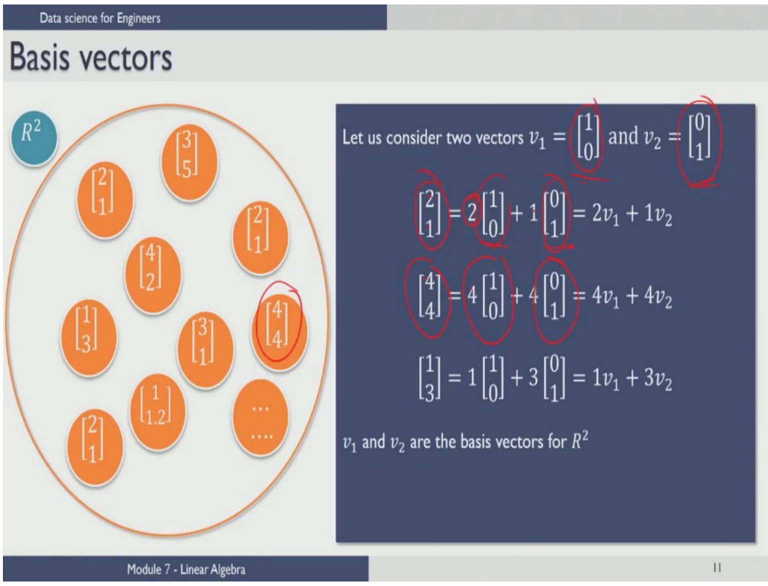
(Refer Slide Time: 09:40)



Now, take the same 2 vectors, which are orthogonal to each other and you know that, when I take a dot product between these 2 vectors it is going to go to 0. If I also impose the condition, that I want each of these vectors to have unit magnitude then what I could possibly do? Is I could take this vector and then divide this vector by the magnitude of this vector.

So, this is going to be root of one squared + - 2 whole squared + 4 squared. Similarly, I can take this vector and divide this vector by the magnitude of the same vector, which is going to be root of 2 squared + 5 squared + 2 squared. Now, these 2 are unit vectors, because the magnitudes are the same and these unit vectors also turn out to be orthogonal to each other, the orthogonal property is not going to be lost, because these are scalar constants. So, while you take $v_1^T v_2$ or $v_2^T v_1$, it will still turn out to be 0. So, these vectors will still be orthogonal to each other. However now individually, they also have unit magnitude such vectors are called are orthonormal vectors, that we have defined here. Notice that all orthonormal vectors are orthogonal by definition.

(Refer Slide Time: 11:14)



Now, we are going to come to the next interesting concept that we would need in data science quite a bit and I am going to explain this concept through very, very simple examples. This can also be very formally defined, what I am going to do is, I am going to try and explain this in a very simple fashion. So that you understand what this means and I also want to give a context, in terms of why these are some things that we are interested in looking at from a data science viewpoint.

So, we are going to introduce the notion of basis vectors. So, the idea here is the following, let us take R squared which basically means that, we are looking at vectors in 2 dimensions. So, I could come up with many many vectors, right? So, there will be infinite number of vectors, which will be in 2 dimensions. So, this is like saying, if I take a 2-dimensional space how many points can I get? So, I can get infinite number of points. Which is what has been represented here.

So, I have put in some vectors and then these dots represent that, there are infinite number of such vectors in this space. Now, we might be interested in understanding, something more general than just saying that there are infinite number of vectors here. So, what we are interested in is, if we can represent all of these vectors using some basic elements and then some combination of these basic elements, is what we are interested in.

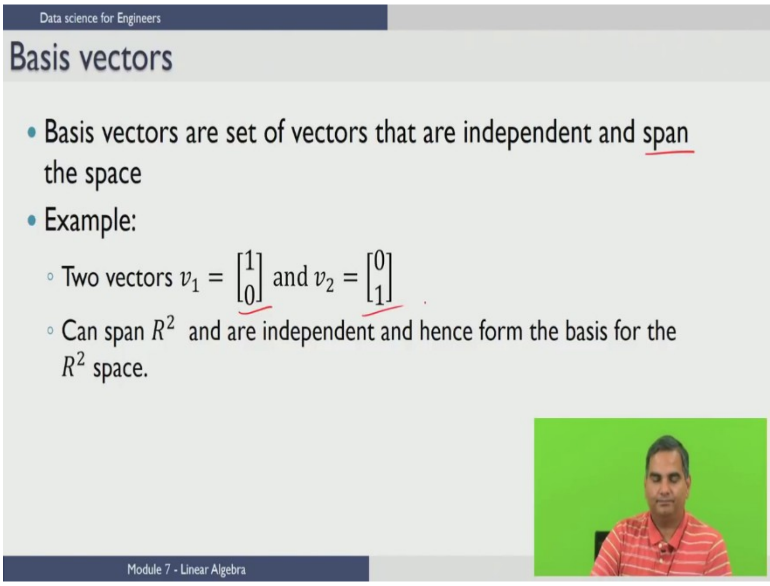
Now, let us consider 2 vectors for example, v_1 1 0 and v_2 0 1. Now, if you take any vector that I have here, let us say take 2 1, I can write 2 1 as some linear combination, of this vector + this vector. Similarly, take 4 4, I can write 4 4 as a linear combination of this vector + this vector and that would be true for any vector that you have in this space.

So, in some sense what we say is that, these 2 vectors characterize the space or they form a basis for the space and any vector in this space can simply be written as a linear combination of these 2 vectors. Now you notice, the linear combinations are actually the numbers themselves. So, for example, if I want this to be written as a linear combination of 1 0 1 0 1, the linear combination the scalar multiples are 2 which is this, and 1 which is this similarly 4 here 4 here and so on.

So, the key point being, while we have infinite number of vectors here, they can all be generated as a linear combination of just 2 vectors and we have shown here, these 2 vectors as 1 0 1 0 1. Now, these 2 vectors are called the basis for the whole space, if I can write every vector in the space as a linear combination of these vectors and these vectors are independent of each of them.

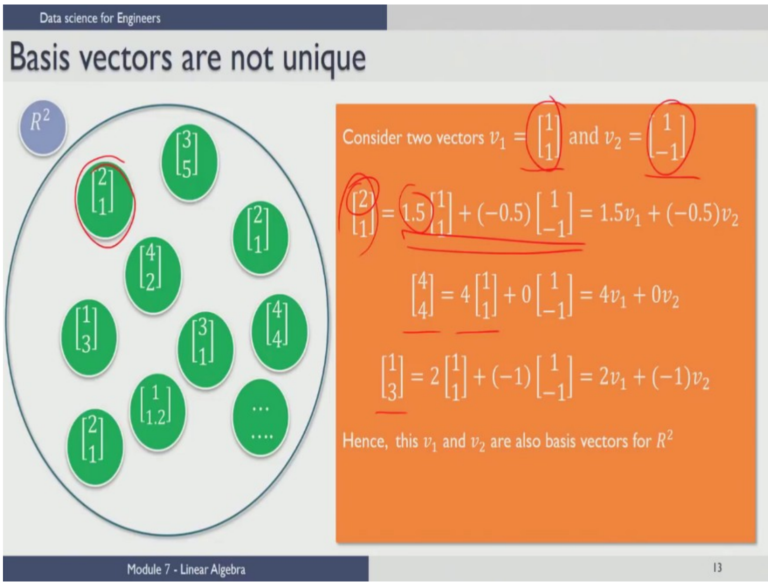
Then we call them as a basis for the space. So, why do you want these vectors to be independent of each other? We want these vectors to be independent of each other, because we want every vector, that is in the basis to generate unique information. If they become dependent on each other, then this vector is not going to bring in anything unique. So, basis has 2 properties, every vector in the basis should some bring something unique, and these vectors in the basis should be enough, to characterize the whole space, in other words the vector should be complete.

(Refer Slide Time: 15:07)



So, this we can formally say as the following, basis vectors for any space are a set of vectors that are independent and span the space and the word span basically means that, any vector in that space, I can write as a linear combination of the basis vectors. So, the previous example, we saw that the 2 vectors v_1 1 0 and v_2 0 1, can span the whole R squared and you can clearly see that they are independent of each other, because no multiple scalar multiple of this will be able to give you this vector .

(Refer Slide Time: 15:49)



So, the next question that immediately pops up in ones head is, if I have a basis vector, are they unique? Now it turns out these basis vectors are not unique, you can find many many sets of a basis vectors, all of which would be equivalent. The only conditions are that they have to be independent and should span the space. So, take the same example and let us consider 2 other vectors, which are independent.

So, the same example as before, where we had used 2 basis vectors 1 0 and 0 1, I am going to replace them by 1 1 and 1 - 1. Now, the first thing that we have to check is, if these vectors are linearly independent or not and that is very easy to verify. If I multiply this vector by any scalar, I will never be able to get this vector. So, for example, if I multiply this by - 1 I will get - 1 and - 1, but not 1 - 1. So, these 2 are linearly independent of each other.

Now, let us take the same vectors and then see what happens. So, remember we represented 2 1 in the previous case, as 2 times 1 0 + 1 times 0 1. Now, let us see whether I can represent this 2 1 as a linear combination of 1 1 and 1 - 1. So, if you look at this, this is the linear combination notice; however, because of the way I have chosen these vectors, these numbers are not the same as this.

So, in the previous case when we use 1 0 on 0 1, we said this can be written as 2 times 1 0 + 1 times 0 1; however, the numbers have changed now, nonetheless I can write this as a linear combination of these 2 basis vectors.

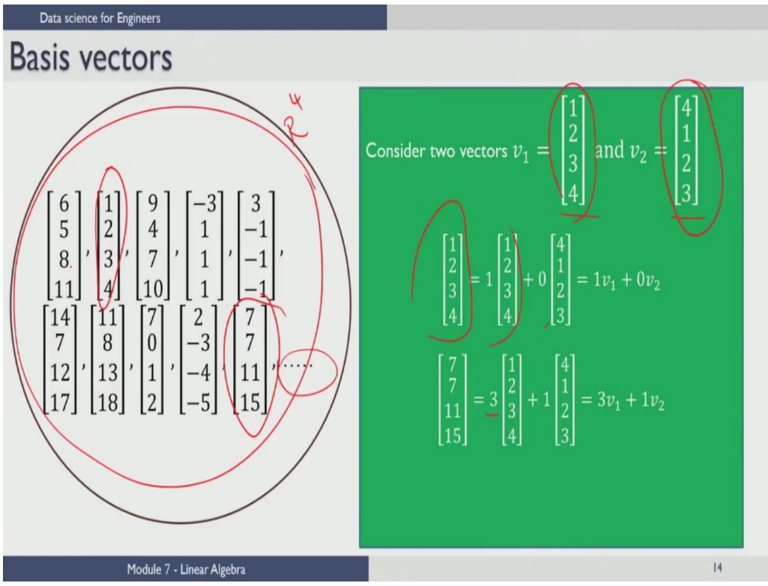
Let us take this 4 4 as an example. So, that can be written as an interesting linear combination, which is 4 times 1 1 + 0 times 1 - 1 right? So, that will give you 4 4 similarly 1 3 can be written as, 2 times 1 1 + - 1 times 1 - 1. So, this is another linear combination of the same basis vectors.

So, the key point that I want to make here is that, the basis vectors are not unique there are many ways in which you can de ne the basis vectors; however, they all share the same property that, if I have a set of vectors which I call as a basis vector, those vectors have to be independent of each other and they should span the whole space and whether you to take 0 1 1 0 and call it a basis set or you take 1 1 and 1 - 1 and call the basis set, both are all right and you can see that, in each case the vectors are independent of each other and they span the whole space.

An interesting thing to note here though is that, I cannot have 2 basis sets which have different number of vectors, what I mean here is in the previous example though the basis vectors were 1 0 and 0 1, there were only 2 vectors. Similarly, in this case the basis vectors are 1 1 and 1 - 1.

However, there are still only 2 vectors. So, while you could have many sets of basis vectors, all of them being equivalent, the number of vectors in each set will be the same. They cannot be different and this is easy to see. I am not going to formally show this, but this is something that you should keep in mind, in other words for the same space you cannot have 2 basis sets - one with n, vectors other one with m vectors - that is not possible. So, if it is a basis set for the same space, the number of vectors in each set should be the same. Now, I do not want you to think that the basis set will always have to be the number of elements in the vector.

(Refer Slide Time: 19:54)



So, to give you another example, we have generated this data in a particular fashion. Consider now this set of vectors right? There are in nite number of vectors here and we will say all of these vectors are in space R 4 , which basically means that there are 4 components in each of these vectors.

Now, what we want to ask is, what is the basis set for these kinds of vectors? Now when I do this here, the assumption is the extra vectors that I keep generating, the infinite number of them, all follow certain pattern that these vectors are also following and we will see what that pattern is. So, what we can do is, we can take, let us say 2 vectors here, in this case this is how this example has been constructed to illustrate an important idea. Let us take this vectors v_1 which is 1 2 3 4 2 which is 4 1 2 3 and let us take some vector here, in this set let us take this vector here, and then see what happens, when I try to write it as a linear combination of these 2 vectors.

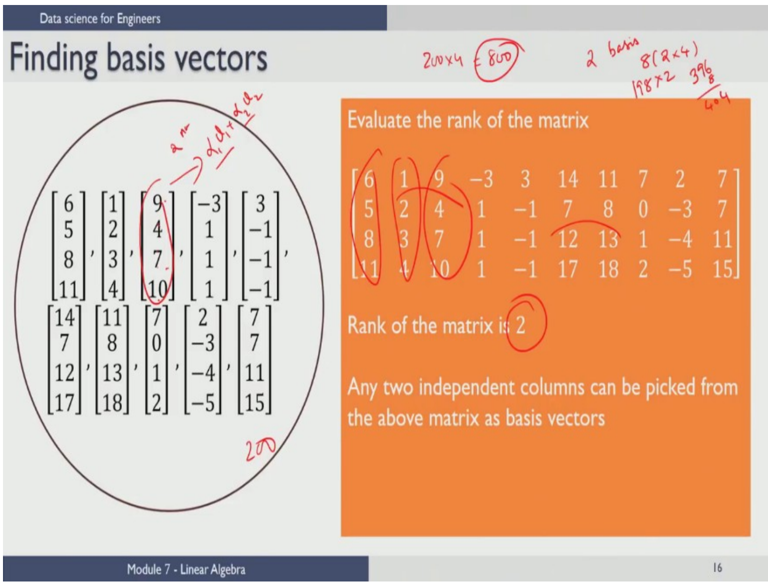
So, I can see that if I take this I can write it as 1 times this + 0 times the second vector. So, that is one linear combination, now let us take some other vector here. So, let us say for example, we have taken this vector 7 7 11 15, we can see that that can be written as a linear combination of 3 times the first vector + 1 times the second vector and so on.

Now, you could do this exercise for each one of these vectors and you will be able to see, because of the way we have constructed these vectors, you will be able to see that each one of these vectors, I can write as a linear combination of v_1 and v_2. So, what this basically says is the following, it says that, though I have 4 components in each of these vectors, that is, all of these vectors are in R4, because of the way in which these vectors have been generated, they do not need 4 basis vectors to explain them, all of these vectors have been derived as a linear combination of just 2 basis vectors, which are given here and here.

So, in other words all of these vectors would occupy A_2 dimensional, what we call as a subspace in R 4 right? So, if you take every vector in R 4, without leaving out anything then, you would need 4 basis vectors to explain all of them. However, these vectors have been picked in such a way, that they are only linear combination of these 2 vectors. So, I just need 2 vectors to represent all of this. So, I say that, all of these vectors fall in a 2 dimensional subspace in R4.

So, this is an important concept of subspace, which is very, very important for us from a data science viewpoint and I am going to explain to you why. We are interested in things like this, from a data science viewpoint. Now, the next question that we might ask is the following.

(Refer Slide Time: 23:30)



So, this is the same as the previous slide, except that I have removed the dot dot dot. So, the way to think about this it is let us say there is some data generation process, which is generating vectors like this, and the dot dot dots that I have left out, will also be generated in the same fashion, because those are also vectors that are being generated by the same data generation process.

So, I have certain data generation process and I am generating samples from that and I have done let us say 10 experiments. So, I have got these 10 samples and the other dots will be similar, now what I want to know is if you give me these, vectors in R4, how many basis vectors do I need to represent them? In the previous slide I had already shown you what the basis vectors are and then shown how I could generate many many linear combinations of just 2 in R4 to get a subspace. I am looking at an inverse problem here, where I do not know what are the vectors that are generating these samples, nonetheless I have got enough samples.

Let us say 10 and if I were to continue this experiment and if it was the same data generation process, I might get 20 samples 30 samples and so on; however, what I want to know is with these 10 samples, how do I find the basis vectors? So, we are going to use concepts that we have learned before to do this. If we were to stack all of these vectors in a matrix like this.

So, this is a first vector here, from here second vector and so on all the way up to the last vector and I say I have so many vectors, how many fundamental vectors do I need to represent all of these as linear combinations? It is a question that I am asking. The answer is straightforward this is something that we have already seen before, if you identify the rank of this matrix it will give you the number of linearly independent columns.

So, what that basically means is, if I get a certain rank for this matrix, then it tells me there are only so many linearly independent columns and every other column, can be written as a linear combination of those independent columns. So, while I have many many columns here, 1 2 all the way up to 10. The rank of the matrix will tell me, how many are fundamental to explaining all of these columns, and how many columns do I need.

So that I can generate the remaining columns as a linear combination of these columns, and as I have been mentioning again, if the data generation process remains the same as I add more and more columns to these, they will also be linear combinations of the columns that we identify here. So, when we go ahead and try to find the rank of this matrix, the rank of the matrix will turn out to be 2 and it will turn out to be 2 because, of the way we have generated this data.

Now, if you had generated these vectors in such a way that they are a linear combination of 3 vectors, then the rank of the matrix would have been 3. If you had generated these vectors in such a manner, that they are linear combinations of 4 linearly independent vectors, then the rank of the matrix would have been 4, but that would be the maximum rank of the matrix, because in R 4 you would not need more than 4 linearly independent vectors to represent all the vectors.

So, the maximum rank can be 4, the rank could be 1 2 or 3. If it is 1 then I have only 1 basis vector, if there are 2 there are 2 basis vectors 3 there are 3 basis vectors and so on. In this case since the rank of the matrix turns out to be 2, there are only 2 column vectors that I need to represent every column in this matrix. So, the basis set has size 2, is something that we have determined. The next question is the basis set is size 2, what are the actual vectors? What we can do is, we can pick any 2 linearly independent columns here and then those could be the basis vectors.

So, for example, I could choose this and this and say, this is the basis vector for all of these columns or I could choose this and this and this or this and this and so on. So, I can choose any 2 columns, as long as they are linearly independent of each other and this is something that we know, from what we have learned before, because we already know that the basis vectors need not be unique. So, I pick any 2 linearly independent columns that represents this data. Now, let me take a minute to explain why this is important from a data science viewpoint. I will just show you some numbers. Supposing, I have let us say 200 such samples and I want to store these 200 samples since each sample has 4 numbers, I would be storing 200 times 4 which is 8 numbers.

Now, let us assume we do the same exercise for these 200 samples and then we find that, we have only 2 basis vectors, which are going to be 2 vectors out of this set. What I could do is, I could store these 2 basis vectors that, would be 8 numbers which is 2 by 4 and for the remaining 198 samples, instead of storing all the samples and all the numbers in each of these samples, what I could do is for each sample I could just store 2 numbers right?

So, for example, if you take this sample, instead of storing all the 4 numbers, I could just store 2 numbers, which are the linear combinations that I am going to use to construct this. So, for example, since I have 2 basis vectors here, there is going to be some number $\alpha 1$ times the basis vector, + $\alpha 2$ times the second basis vector, which will give me this sample right?

So, instead of storing these 4 numbers, I could simply store these 2 constants and since I already have stored the basis vectors, whenever I want to reconstruct this, I can simply take the first constant and multiply v 1 + the second constant multiply v 2 and I will get this number. So, I store 2 basis vectors which gives me 8 numbers and then for the remaining 198 samples, I simply store 2 constants. So, this would give me 396 + 8 404 numbers stored. I will be able to reconstruct the whole data set.

So, compare that with 800. So, I have half reduction in number. So, when you have vectors in multiple dimensions, let us say you have vectors in 10 dimensions 20 dimensions and the number of basis vectors, are much lower than those numbers. So, for example, if you have $A_3$0-dimensional vector and the basis vectors are just 3, then you can see the kind of reduction that you will get in terms of data storage. So, this is one viewpoint from data science. Why? It is very important to understand and characterize the data in terms of what fundamentally characterizes the data. So that you can store less, we can do smarter computations and there are many other reasons why we will want to do this, you can identify this basis to identify a model between this data, you can identify a basis to do noise reduction in the data and so on.

So, all of those viewpoints we will talk about as we go forward, with this data science course. In the next lecture, we will continue and then try and understand how we can use these concepts. The notion of basis vectors, the notion of orthogonality to understand concepts such as projections, hyper planes, half spaces and so on, which all are critical from a data science viewpoint. So, I will pick up from here in the next lecture

Thank you.