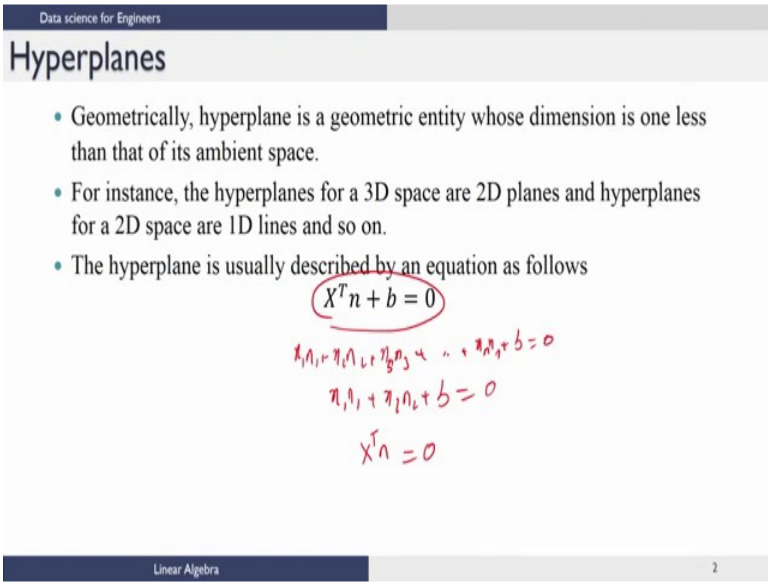
Data Science for Engineers
Prof. Raghunathan Rengasamy
Department of Computer Science and Engineering
Indian Institute of Technology, Madras

Lecture- 17
Linear Algebra - Distance, Hyperplanes and Halfspaces, Eigenvalues, Eigenvectors

We will continue with our lectures on linear algebra for data science. Today I will talk about hyper planes, half spaces and eigenvalues, eigenvectors and so on.

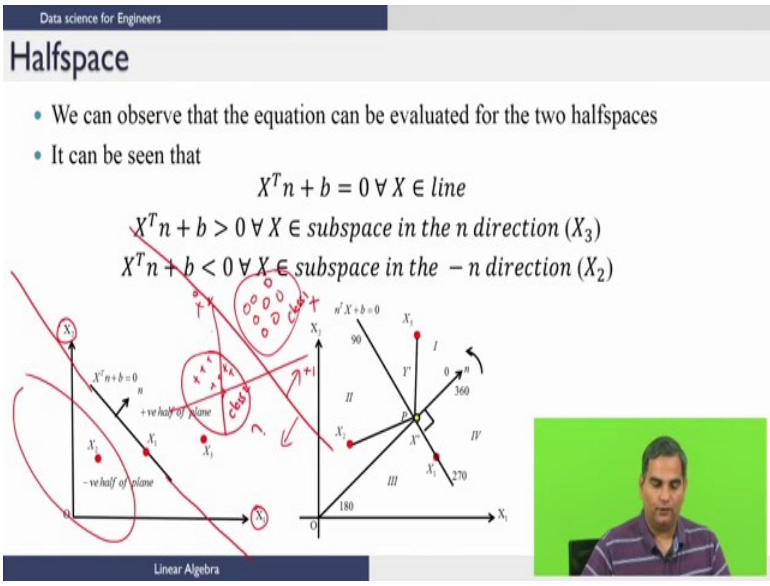
(Refer Slide Time: 00:27)



Let us start this lecture with hyper planes. Geometrically hyper plane is a geometric entity whose dimension is 1 less that than that of its ambient space. So, what this means is, the following. For example, if you take the 3D space then hyper plane is a geometric entity which is 1 dimension less. So, its going to be 2 dimensions and a 2 dimensional entity in a 3D space would be a plane.

Now if you take 2 dimensions, then 1 dimension less would be a single dimensional geometric entity, which would be a line and so on. The hyper plane is usually described by an equation as follows, if I expand this out for n variables. So, I will get something like X_1 n_1 + X_2 n_2 + X_3 n3 and so on Xn n_n + b = 0 in just two dimensions, you will see that this is X_1 n_1 + X_2 n_2 + b = 0 which is an equation of line. We have seen before the idea of subspaces. Hyperplanes in general are not subspaces, however, if we have hyper planes of the form X transpose n = 0; that is if the plane goes through the origin, then an hyper plane also becomes a subspace.

(Refer Slide Time: 02:03)



Now that we have described what a hyper plane is, let me move on to the concept of half space to explain the concept of half space. I am going to look at this 2 dimensional picture on the left hand side of the screen. So, here we have a 2 dimensional space in X_1 and X_2 and as we have discussed before an equation in two dimensions would be a line which would be a hyperplane. So, the equation to the line is written as $X^T n + b = 0$.So, for, in this two dimensions we could write this line as; for example, $X_1 n_1 + X_2 n_2 + b = 0$, while I have drawn this line only for part of this picture. In reality this line would extend all the way on both sides.

Now, you notice the following. You see when I extend this line all the way on both sides, then this whole two dimensional space is broken into two spaces, one on this side of a line and the other one on this side of a line. Now these two spaces are what are called the half spaces. Now the question that we have is the following.

If there are points on one half space and points on the other half space, is there some characteristic that separates them? For example, can I do some computations for all the points on one half space and get some value and some computation for all the points on the other half space and get some value and use that to make some decisions and that is a reason why we are interested in this half spaces from a data science viewpoint.

So, this question is of importance in a particular kind of problem called a classification problem. Let me explain what that means. In fact, we are going to look at a very specific classification problem called binary classification problem. So, let us assume that I have, let us say in two dimensional space a data belonging to two classes.

For example, let us say I have data belonging to class one like this, and I call it class one and then I have data belonging to class two is something like this, call it class 2. So, this classes could be anything. So, for example, this could be a group of people who like South Indian restaurants and this could be a group of people, who do not like South Indian restaurants, and the coordinates X_1 and X_2 could be some way of characterizing people in terms of some attributes of these folks. Let us say we have taken a survey to say whether they like South Indian food or do not like South Indian food.

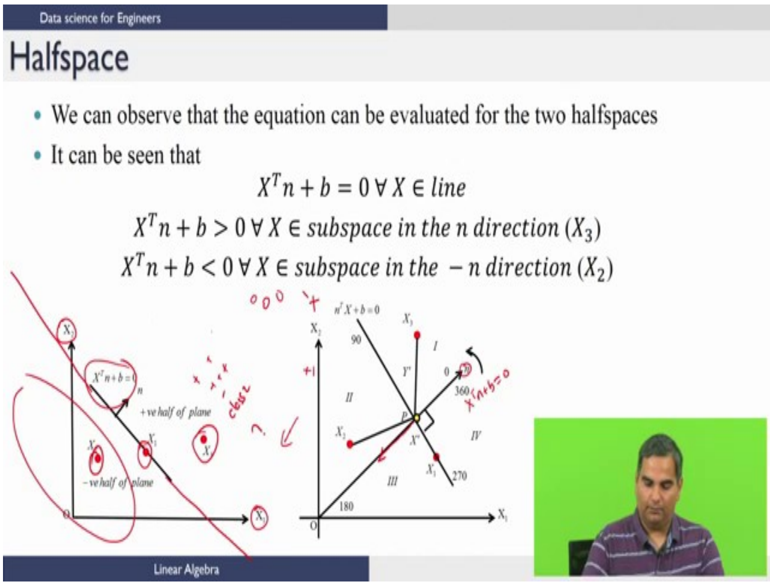
Now what we want to do is, if I give you the attributes of a new person, let us say that attribute falls here and then I ask you this question as to would this person like South Indian food or not like South Indian food and the answer would most likely be that this person will not like South Indian food, because this data point is very close to class 2.

Whereas if I gave you another point here for example, then you would come to the conclusion, this person is likely to like South Indian food. So, what we want to do is, we want to be able to evaluate cases like this. So, we want to somehow come up with a discriminating function between these two classes. So, one way to do that would be something like this; draw a line between these two classes and then say, if there is some characteristic that holds for this side of the line, which is what we called as a half space here. And if there is some characteristic that holds to this side of the line then we could use that characteristic as a discriminant function for doing this binary classification problem. So, that is the data science interest in understanding this topic in linear algebra.

Now let us proceed to see how we do this through some simple geometric concept. Let us go back to this picture and then ask the question as to how do I determine which side of a half plane or a half space, which half space does a point lie in. So, to understand this, what we are going to do is, we are going to take three points as shown here X_2 X_1 and X_3 and ask the question as to how do I distinguish whether the point is on a line or to one half space or the other. So, the way we are going to do this, is the following. We are going to first look at this in little more detail and we know that when I write an equation of the form $X^Tn + b = 0$, n is normal to this line, is something that we have already described.

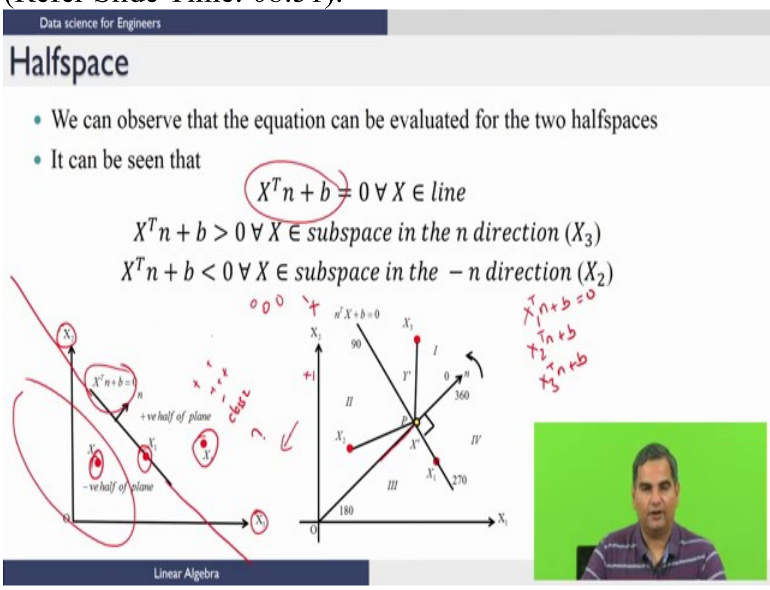
However there is an important point to note. Here the normal could be defined in two ways, one is the normal is in this direction, the other thing to do is to just take the opposite direction and then define a normal in this fashion also. So, its important to know in which side normal is defined to understand this. For example, if I say this is a normal for an equation which is $X^T n + b = 0$.

(Refer Slide Time: 08:16)



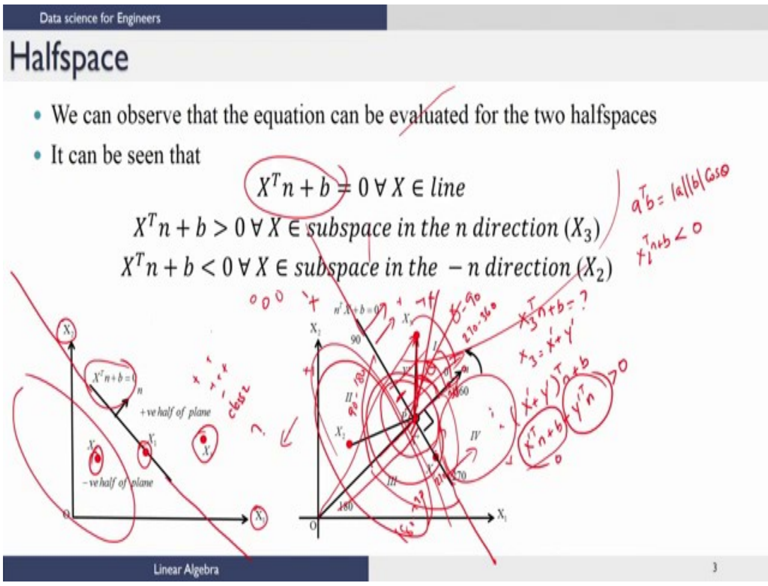
If I simply multiply this equation by - 1, then I am defining a normal to the other side. So, this is an important point to remember. Now what we want to know is, where do this points X_1, X_2, X_3 lie to do this. What we are going to do is, we are going to evaluate a discriminant function or a function which is basically the equation of the line. So, what we want to do is.

(Refer Slide Time: 08:51).



We want to understand what this will be, what this will be and what this will be. Now, when we look at point X_1 we know that the point lies on the line. So, this is going to be 0. So, this is straightforward. What we are interested in, is what happens to this quantity for X_3 and X_2, and is there some way in which we can say that every point to one side of the line will have the same characteristic and every other point on the other side of the line will have a different characteristic. So, to do this, let us first look at $X_3^T n + b$ and then see what happens.

(Refer Slide Time: 09:44)



So, I want to know what this is. Notice in this picture I have defined a new point X on the line and then I have another vector which goes from X′ to X_3. Now X_3 is the vector that goes from here to here. From vector addition we know that I can write X_3 as X’, this + this X’ + Y’. So, what I am going to do is, I am going to simply substitute this into the equation and then see what happens. So, I am going to have X’ + Y’^{T} n + b.

This is what I want to evaluate. This will become X’^{T} n + b + Y’ ^{T} n. All I have done is, I moved b closer to this term to show you something. Now notice what happens to this term right here, since X’ is on the line and the equation of line is X ^{T} n + b this has to go to 0. So, when we compute X_3 ^{T} n + b, we are simply left with this term right here. And if you notice this term you would see that this is a dot product between this vector and this vector, and the most important thing to note here is the following, as long as the point lies to this side, this side of the line then you would see whatever point you take, the angle between that point and the normal would be in the following ranges.

So, you take any point this side or this side. So, the angle between the normal and that point is going to be the following. So, supposing we look at this and then say; I am going to do this angle in this direction right. So, what you are going to notice is the following. If the point is between these two, then I am going to have a positive θ angle. Now the way you do this is the following.

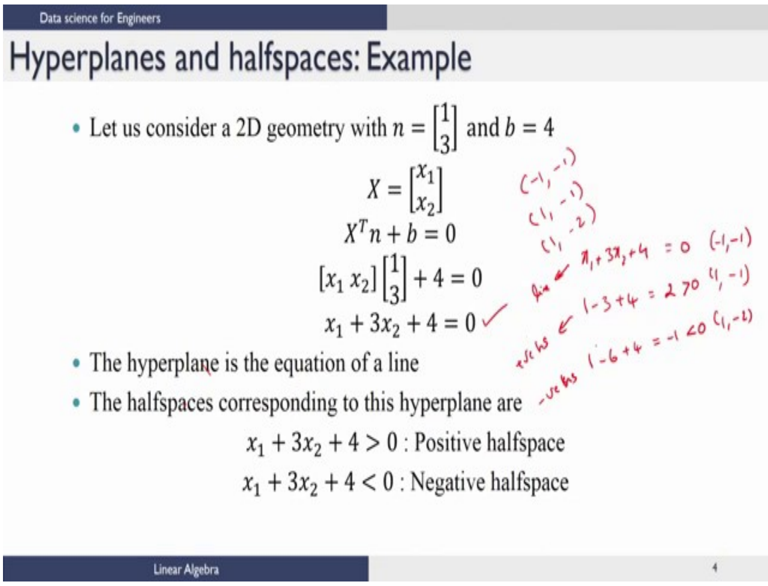
So, you go like this. So, for this quadrant if you start with 0 here for this quadrant the angle is going to be between 0 and 90, and for this quadrant the angle is going to be between 270 and 360. So, if a point is this side, the angle between this vector and this normal is going to be between 0 and 90. And if the point is in this side the angle is going to be between 270 and 360. We also know that when I have dot products A^T b, I can also write this as magnitude of a magnitude of b cos θ, where θ is the angle between these two vectors.

So, we will look at all the points up to here. So, whatever is a point you have these angles and all of these angles are between 0 and 90. So, for any point between here and here in this whole space you are going to get a b, some angle between 0 and 90, and we know from our high school rule, all silver teacups cos θ will always be positive. So, A^T b is going to be positive; that means, this is going to be positive. Now when you get two points here then the angles are going to be between 270 and 360 which is in the fourth quadrant. Again using the same rule all silver teacups the fourth quadrant is c cos. So, cos is going to be positive. So, again you have A^T b being positive.

So, irrespective of where the point is to this side of the line, when I take this $X_3{}^T n + b$, I am always going to get a positive value. Now by similar argument you can say for any point on the other side or the other half space, the angles are going to be between 90 to 180 here and 180 to 270 and as we know cos θ for angles between 90 to 270 is negative.

So, any point on this side of the line or the half space the computation $X_2{}^T n + b$ is going to be less than 0. So, this is an important idea that that I would like you to understand. So, what this basically says is the following. If you were to simply take any point that I give you and then I evaluate $X^T n + b$, if that point is on the side of the normal half space then $X^T + b$ will be positive, and if its on the half space in the opposite side then its going to be negative. And I already told you how this is important from a data science viewpoint.

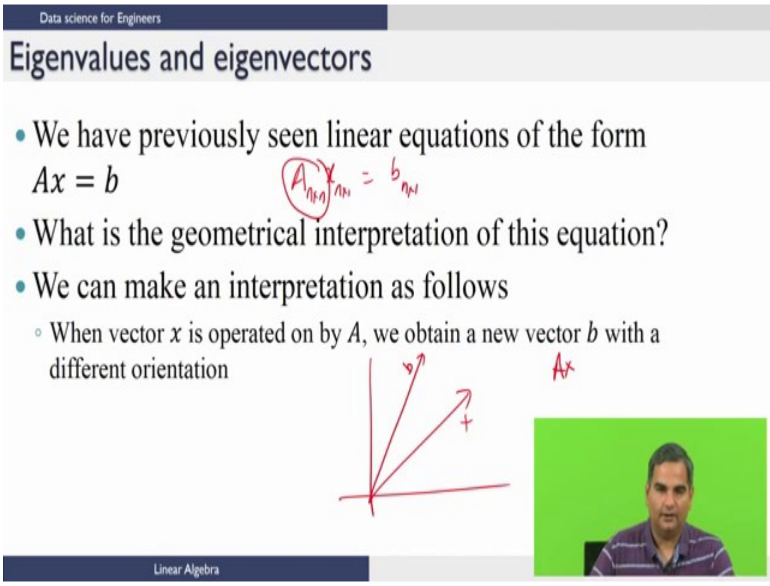
(Refer Slide Time: 15:41)



So, let us consider simple 2D geometry and then let us take n as 1 3 and b as 4. So, this would give me this equation for $X^T n + b$. Now let us say I take a point on three points for example. So, let me consider - 1 - 1 as one point, let us also consider 1 - 1 as another point and let us consider 1 - 2 as another point and then see what happens. So, when I take the point - 1 - 1 and I substitute into this x 1 + 3 x 2 + 4. So, it will be - 1 - 3 + 4. So, the point 1 - 1 will lead to - 1 - 1. Sorry will lead to 0. So; that means, the point - 1 - 1 is on the line, when I take the point 1 - 1. So, this is going to be 1 - 3 + 4. So, this is going to be 2, so positive.

So, this is on in the positive half space and when I take the point 1 - 2 then I am going to get 1 - 6 + 4 which is going to be = - 1 less than 0. So, this is in the negative half space. So, this is on the hyperplane of the line, this is on the positive half space and this is on the negative half space. So, that tells you how to look at different points and then decide which side of the hyperplane or which half space these points lie.

(Refer Slide Time: 17:35)

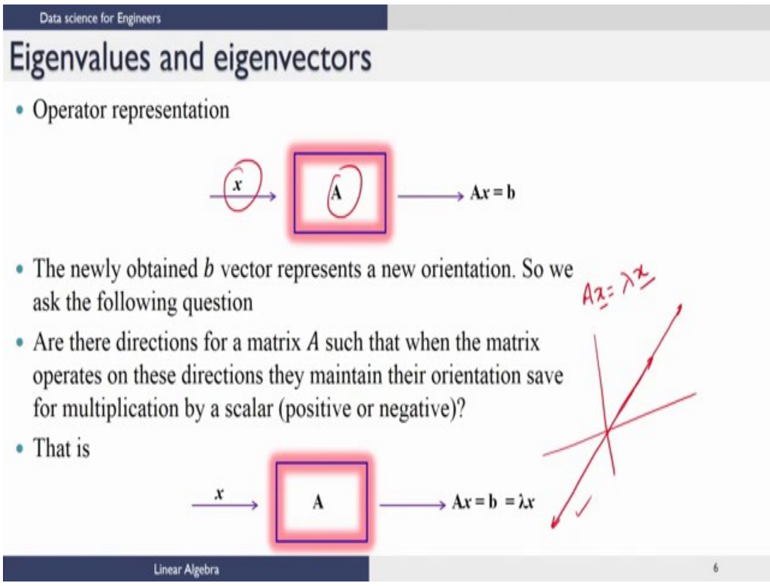


Now, that we have understood hyper planes and half spaces we are going to move on to the last linear algebra concept that I am going to teach in this module on linear algebra for data sciences. And once we are done with this topic, then we have enough information for us to teach you the various algorithms, commonly used algorithms or the first level algorithms in data science. So, let us look at this idea of eigenvalues and eigenvectors, we have previously seen linear equations of the form Ax = b. We have looked at it both algebraically and geometrically, we have spent quite a bit of time on looking at these equations algebraically. We talked about when these equations are solvable when there will be infinite number of solutions, how do we address all of those cases in a unified fashion and so on.

Now what we are going to do is, now that we know both vectors and so on, we are going to look at a slightly geometrical interpretation for this equation again and then explain the idea of eigenvalues and eigenvectors and then con$_n$ect the notion of these eigenvalues and eigenvectors with the column space, null space and so on that we have seen before. So, this is very important, because these ideas are used quite a bit in data compression, noise removal, model building and so on. We will start saying I have this Ax = b and a is an n by n matrix x is n by 1 and b is n by 1. So, this is the kind of system that we are looking at.

So, we are going only look at square matrices n by n. Now you can think of this as n equations and n variables. There is also another interpretation you can give for this which is the following, supposing I have a vector x something like this and if I operate A on this. So, by operating, I mean we define an operation as pre multiplying this vector by A. So, let us say I operate A on this vector which is Ax then I notice from this equation I get b, which is basically some other new direction that I have. So, you can think about this as the following I can think about this as a equation, which tells me that when I operate A on x then I get a new vector b which is in a different direction from x. So, this is a very simple interpretation of this equation Ax = b, which is what is written here x, I send it through a and I define sending it through a as pre multiplying by A; so A times x equal b.

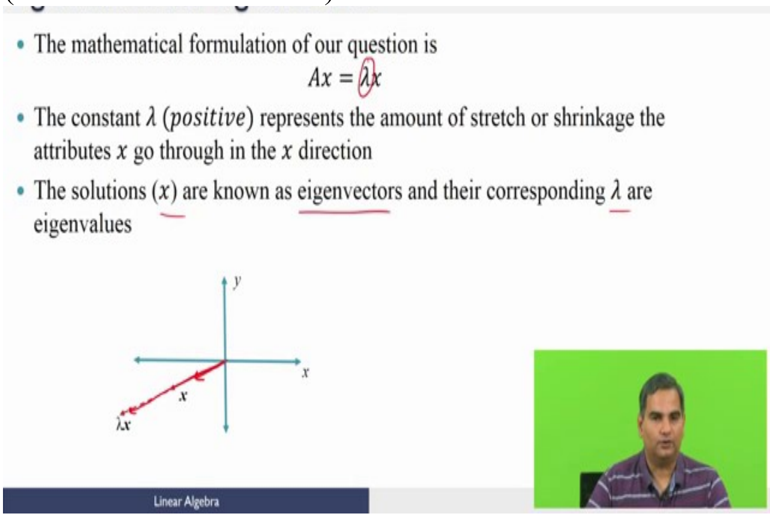
(Refer Slide Time: 20:45)



Now, that we have this interpretation, we ask the following question for a matrix A. Are there some directions which when you operate this A on they do not change their orientation. In other words I want to know if there are x vectors for matrix A such that when I operate A on x I get λ x not b, λ x here, the idea is because this is x, there is no change in orientation safe multiplication by a scalar. Now this multiplication scalar could be positive or negative, in which case we are talking about the following.

So, if this is x. So, when I operate A on x, since its in the same direction, its either this way or this way and if λ is positive, it will be in this direction and if λ is negative, it will be in this direction and so on. So, the question is, would there be directions like this for all kinds of matrices is an interesting question, that you could ask for.

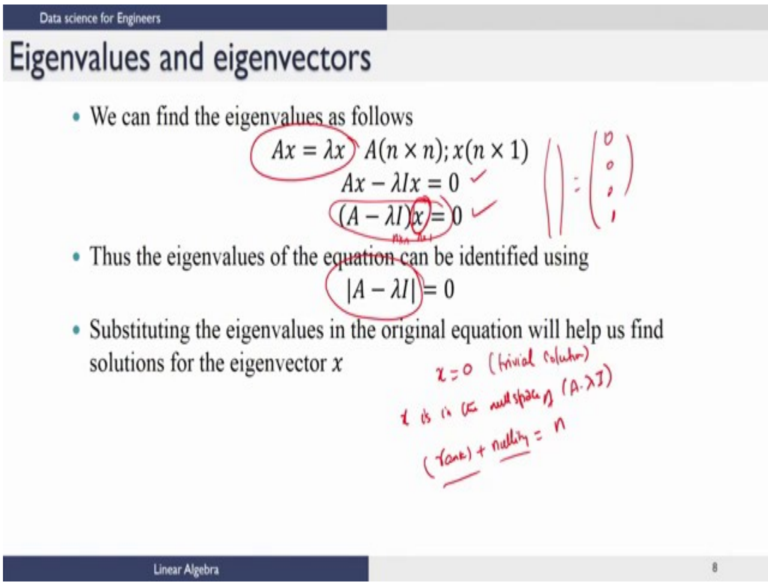
(Refer Slide Time: 22:04)



Now let us focus on λ being positive, λ can be negative also. If λ is positive, then we see this equation and then notice that if λ is less than 1, then basically when I operate A on x the vector actually shrinks. So, if this is x, if λ is less than 1, this will be shrunk like this, and if λ is greater than 1 it will be at a higher magnitude than the original x vector.

Now the question is, for every matrix A, would there be A vectors like this x and what would be the scalar multiple and what is the use of all of this, is something that we should also address at some point as we go through this lecture. Now let me give you some definitions, this x are called eigenvectors and lambdas are called the eigenvalues corresponding to those eigenvectors. So, the questions that we are left with, are how do we find out that every matrix, whether it would have eigenvectors and how do I compute this eigenvectors and eigenvalues.

(Refer Slide Time: 23:26)



So, to compute the eigenvalues we follow this procedure that I am going to outline now. So, the original equation is A x = λ x, what I could do is, I could bring this λ x to this side, and then I get this equation Ax - λ I x = 0. So, this becomes A - λ I times x = 0. Now, notice that this is basically A vector equation, because I have n by n vector this is n by 1. So, I have on the left hand side n by 1.

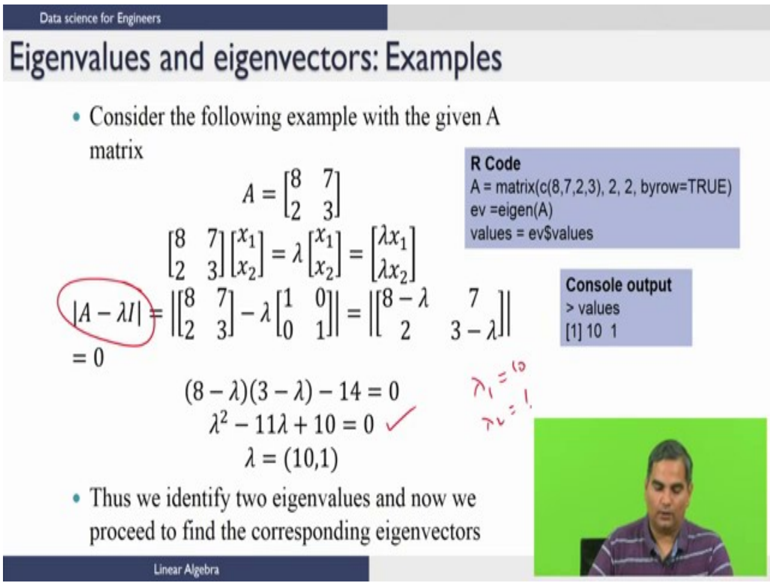
So, I have a vector here and I want 0s here. So, I want to find an x; such that this is true. Now we have everything that we need to solve at this equation. So, what I am going to explain to you is the following. If I want to get an x which is not all 0, notice that if x is all 0, this is a solution right. So, x = 0 is a solution, but we are not interested in this solution, because this is what we call as a trivial solution.

We are not interested in this, we are only interested in solutions that we call as non trivial, at least one of the xs will have to be non 0. Now notice that if this equation is solvable, then x is in the null space of A - λ I matrix. This is something that we have seen before, while we de ne the null space and we also know that the rank nullity theorem says the rank of the matrix + nullity = n which is the number of columns, we are looking at square matrices n by n matrices. Now we know that if there is even one vector x; such that this is 0; that means, then rank of the null space is at least 1, and since the rank of the null space is at least one nullity is at least 1; that means, the rank of the matrix has to be less than n right, it can$_n$ot be n, if this is n nullity is 0, that means, there are no non trivial solutions.

So, if there needs to be a solution for x, then we know that the rank of the matrix A - λ I has to be less than n; that is the matrix A - λ I is not a full rank matrix, and we know that if the matrix is not full rank then the determinant of that matrix has to be 0. So, in summary if we want a non trivial solution for x, then that necessarily means that this determinant A - λ I has to be = 0.

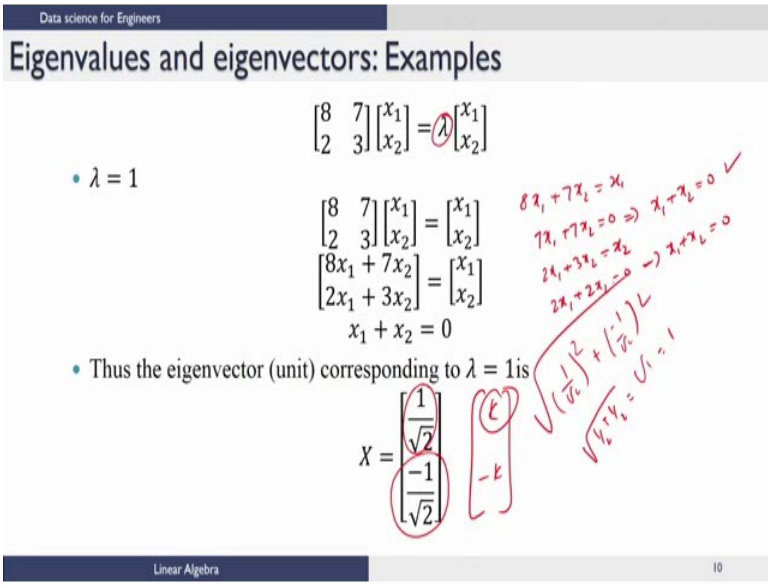
Now once we solve for this equation and compute A λ, then we can go back and then substitute the value of λ here and then we have A - λ I times x = 0, the way we have chosen λ is such that this matrix does not have full rank; that means, there is at least one vector in the null space, and using concepts that we have learned before we can identify this null space vector which would become the eigenvector.

(Refer Slide Time: 27:02)



Let me illustrate this with an example here, let us consider the matrix A which is 8 7 2 3 and let us compute this determinant A - λ I. So, you get the following equation and you get a quadratic equation here. Notice an interesting thing here; if I have an n by n matrix, the determinant in λ would be an nth order polynomial. In this case I have A_2 by 2 matrix. So, the determinant is a λ function which is a quadratic, and if its 3 by 3 it will be cubic and so on. So, this opens out the possibility of a solution to this equation being complex also. So, this is an important point to note here though your original matrix A is real. The solution to your eigenvalue problem could be either real or complex, depending on the polynomial that you end up with. In this case we have chosen this example in such a man$_n$er that I get two real solutions and the real solutions are 10 and 1. So, I can easily see that this equation has solutions 10 and 1. So; that means, I have two eigenvalues λ 1 = 10 and λ 1 = 1. Now how do I go ahead and calculate the eigen vectors corresponding to these eigen values.

(Refer Slide Time: 28:25)

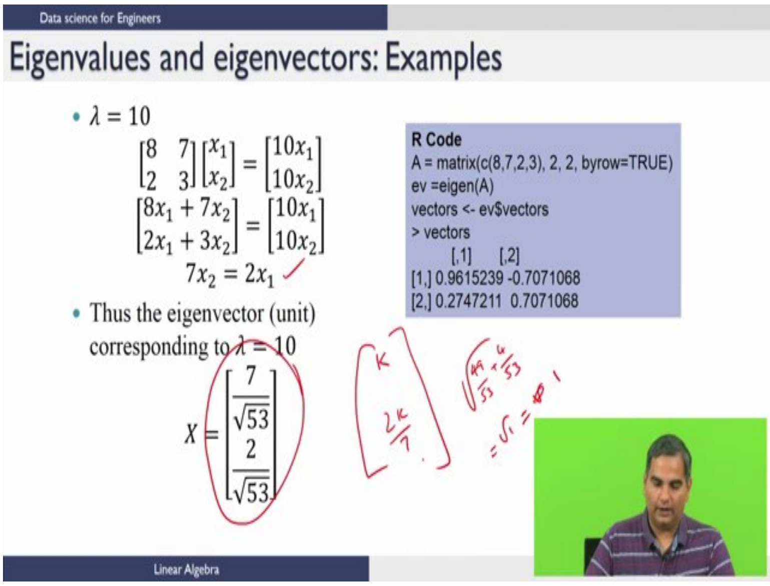


So, let us illustrate this for λ = 1. So, I take this eigenvalue eigen-vector equation now that I know λ = 1, this becomes 8 7 2 3 X_1 x is X_1 + X_2. Now this turns out into these two equations, and if you notice you take the first equation, the first equation is $8X_1 + 7X_2 = x\ 1$.

So, if I take X_1 to this side I get $7X_1 + 7X_2 = 0$, which is the same as $X_1 + x = 0$. If you take the second equation you will see that it is $2\ X_1 + 3X_2 = X_2$ which basically says $2x + 2\ X_2 = 0$, which also is $X_1 + X_2 = 0$. So, both these equations turn out to be the same. Now any solution where X_2 is the negative of X_1 would be a eigenvector, what we do is, the following of all of those solutions.

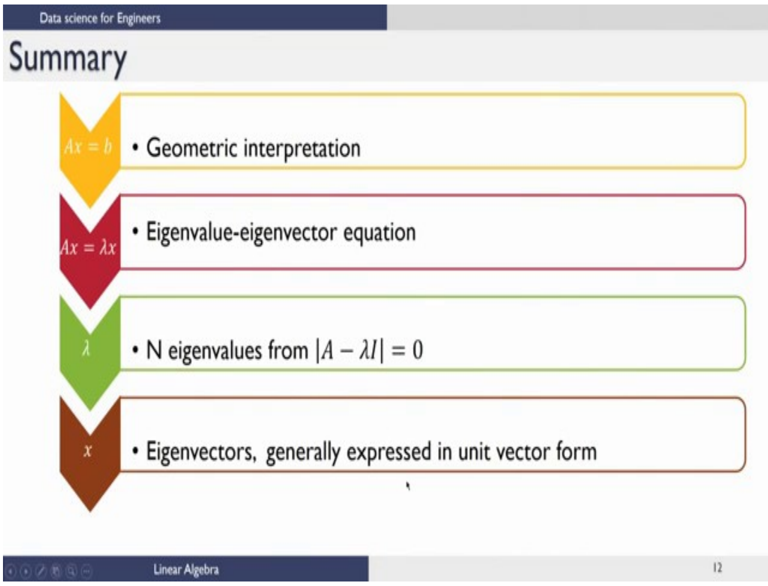
We also make sure that we get an eigenvector which has unit magnitude. So, if you notice here the eigenvector that we get, you notice that X_1, and this is X_2 and you notice that X_2 is - X_1 or X_1 is - X_2, which is what will satisfy this equation and instead of picking any k - k as a solution here, we pick a k in such a way that the magnitude of this vector is 1. So, we know that the magnitude of this vector will be 1 by root 2 whole square + - 1 by root 2 whole square root which will be root of half + half which will be root of 1 = 1. So, that way we make this a specific eigenvector which is unit length. We could do the same thing for λ equal 10, by much the same procedure you will notice that you will get this equation here $7X_2$ is 2x .

(Refer Slide Time: 30:30)



So, basically what you could do is, any vector which is such that if X_1 is kX_2 is 2k by 7 would satisfy this equation; however, what we do is, we choose this k in such a way that the magnitude of the eigenvector is 1. So, in this case 7 by root 53 2 by root 53, if you do the magnitude of this you will see this is going to be root of 49 by 53 + 4 by 53, which will be root of 1 = 1. So, you see that the magnitude is 1 and also this equation is basically satisfied by any eigenvector which is of this form k to k by 7.

(Refer Slide Time: 31:33)



So, in summary for the eigenvalue eigenvector portion of this lecture, we started with A x = b which has a geometric interpretation of A operating on x giving a new vector b. Now if we force this b to be λ x some scalar multiple of x itself, where the scalar multiple could be either positive or negative, we get the eigenvalue eigenvector equation and to calculate the eigenvalue. What we do is, we calculate the determinant $A - \lambda$. I set it to 0 for an n by n matrix, there will be an nth order polynomial that we need to solve which opens out to the possibility of the eigenvalues, being either real or complex. And once we identify the eigenvalues we can get eigenvectors as the null space of $A - \lambda I$ where λ is the corresponding eigenvalue.

In the next lecture what I will do is, I will con$_n$ect this notion of eigenvalues and eigenvectors to things that we have already talked before in terms of column space and null space of matrices and so on. We already saw that the eigenvectors are actually in the null space of $A - \lambda I$, I am going to develop on this idea and then show you other con$_n$ections between eigenvectors and these fundamental subspaces, and I will also allude to how this is a very important problem; that is used in a number of data science algorithms. So, I will see you in the next class.

Thank You.