Artificial Intelligence for Economics

Prof. Dripto Bakshi

Humanities and Social Sciences

Indian Institute of Technology Kharagpur

Week – 07

Lecture - 35

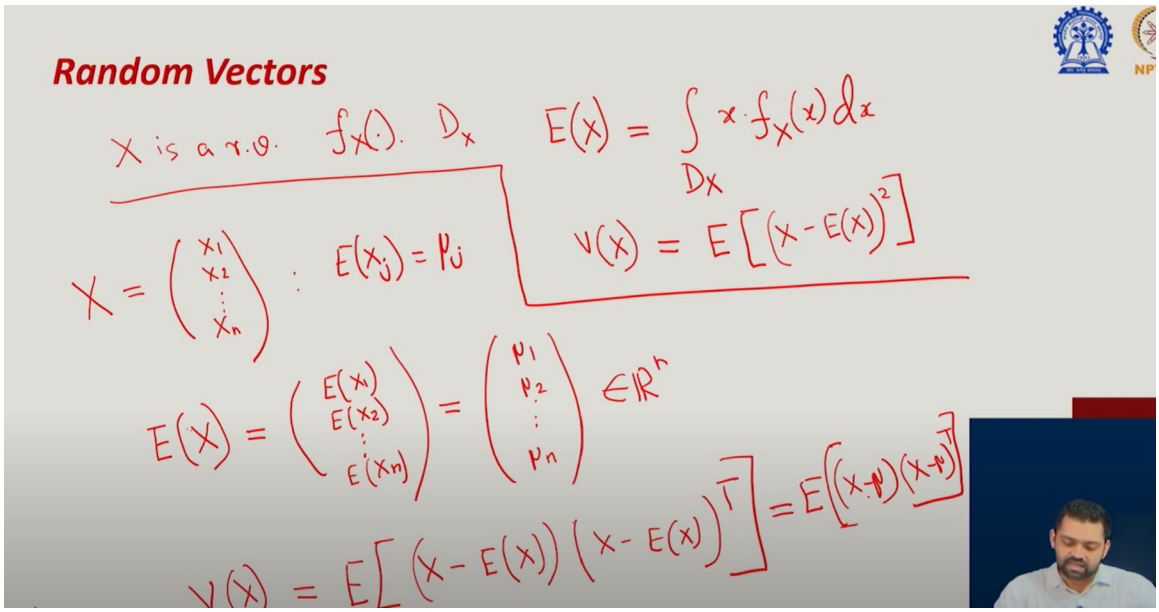
Lecture 35 : Dimensionality Reduction (Principal Component Analysis) – The Math Prerequisites

Welcome to this lecture of artificial intelligence for economics. In this lecture, we will begin with a very important technique for dimensionality reduction, which is principal component analysis. Now, nowadays when you encounter huge data sets, where the number of explanatory variables are huge, then you need to shrink the data and reduce the number of variables. And this is one of those techniques. This is one such technique which leads to, ah, reducing the dimensionality of the data. But before, ah, and this is what we will study for the, for the, for the two lectures, this and the next one.

So before we get into the technique, ah, right away, let's understand the mathematical prerequisites. You just need to be, ah, you need to know a little bit of statistics, some elementary linear algebra, and Lagrange optimization. So, we will first cover the mathematical prerequisites in a short lecture and then we will get into the real technique great. So, let us get started.

The first thing which we need to know is random vectors or vectors of random variables. So, let us say x is a random variable x is a random variable we know that and let us say x has a density function given by capital $F_X(\cdot)$ and let us say x has a domain or a support given by $D_X(\cdot)$. Then we know that a random variable is usually characterized by or the two features of the random variable which we are typically interested in are expectation of x which is simply given by integrating over the domain of x and then we know variance of x is given by or let me write it here, right. let us understand this is this is for a random variable one random variable. But what if we have a vector of random variables like this let us say x is a vector of random variables then how can we find the expectation and variance of a vector of random variables ok.

So, let us understand. expectation of any random variable X_j is given by μ_j let us say. So, what is expectation of a vector of random variables? Well it is simply vector of the



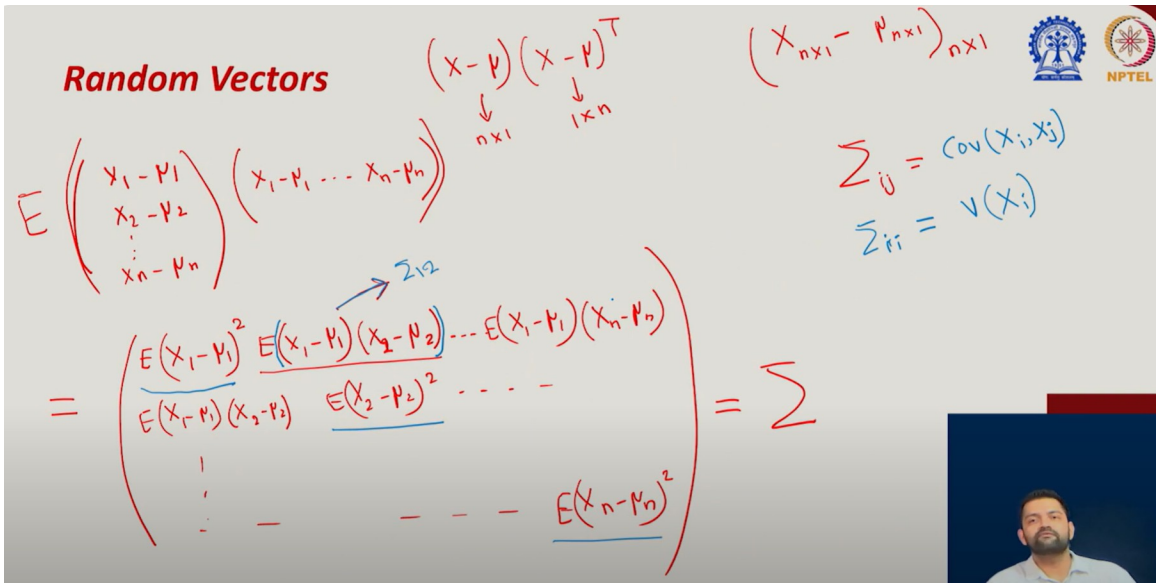
expectations as simple as that. So, this is simply and what would that be $\mu_1, \mu_2, \ldots, \mu_n$ which is nothing but a point in the n dimensional plane right. So, that is expectation of x where x is a random vector is a vector of random variables.

What about variance of x variance of this vector of random variables what will that be given by Well, variance of x is given by this. Take a look at this carefully. So this is $X - E[X]$. Remember, x is a vector, expectation of x is also a vector. Transpose.

Okay? So that's what variance of x is given by. Or if we just write it a little more simply this is $(x-\mu)\cdot(x-\mu)^{tr}$ that is variance of x. Now, you can guess that what will this lead to what is this $(x-\mu)\cdot(x-\mu)^{tr}$ let us see. Let us look at this matrix. x is $n \times 1$ right μ is $n \times 1$.

So, x - μ will be $n \times 1$ vector what about $(x-\mu)^{tr}$ that will be a $1 \times n$ vector. So, if this is a $n \times 1$ vector and this is a $1 \times n$ vector if we multiply them we will get a $n \times n$ matrix. So how will $(x-\mu)\cdot(x-\mu)^{tr}$ look like? Let us see. Let us write this down. So $x - \mu$ is simply this.

And then you have, okay? Now if we multiply these two, what will we get? You will get something like this. That's the first row, the second row will be and so on and so forth and the last element is going to be. So this is how $(x-\mu)\cdot(x-\mu)^{tr}$ looks like. What if we take expectation? remember that is what it was variance of x was expectation of this matrix $(x-\mu)\cdot(x-\mu)^{tr}$ remember just like in a vector if you take expectation of a vector of random variables you get a vector of the expectations of the of the constituent random variables here also if you take expectation of of this matrix it is same as putting expectation on all the constituent terms. So it is simply this, this, this, this, this, this,



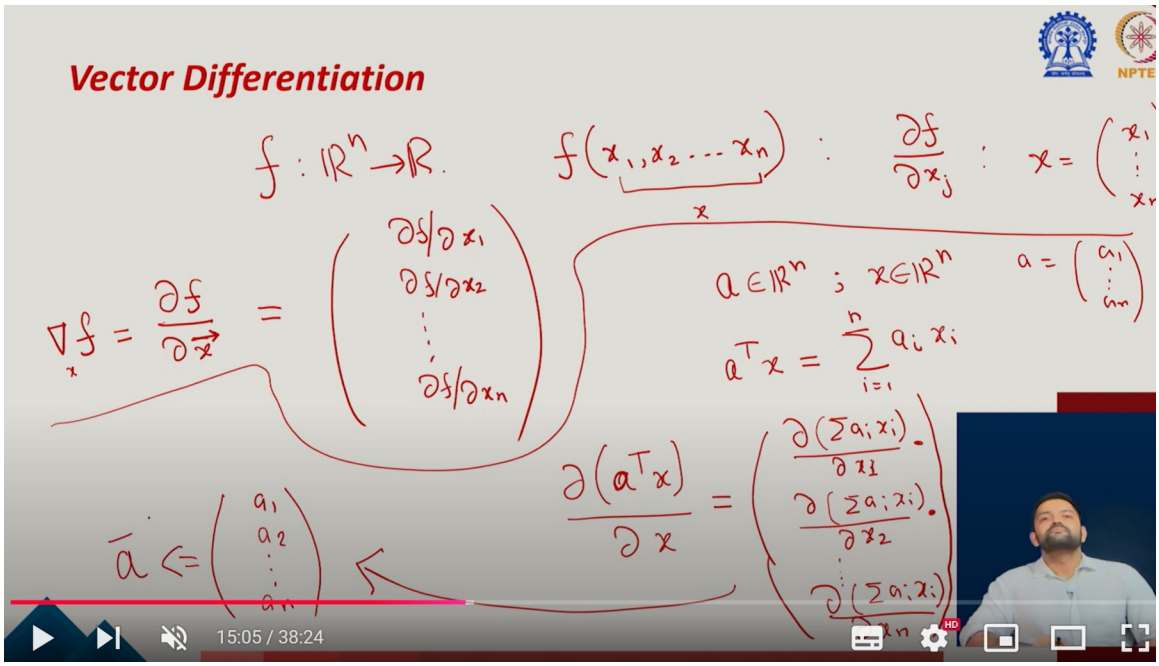
okay.

So let us call this matrix Σ. Now can you guess what will, what is Σ_{ij} where $i \neq j$ let us say. Let us say this is Σ_1, this. Let me change the color of the pen just to make it little more identifiable. So, this is Σ_{12} and what is this? Well, this is covariance between X_1 and X_2, okay.

So, if you have two random variables X_1 and X_2 and their means are μ_1 and μ_2, then that is the definition of covariance between X_1 and X_2. So Σ_{ij} will simply be covariance between X_i and X_j. What about the diagonal terms? The first diagonal term is simply variance of X_1. The second diagonal term is the variance of X_2. So on and so forth.

So Σ_{ij} is simply variance of X_i ok great. So, we have learnt about random vectors how to how to compute expectation and variance of a random vector ok. If we take expectation of a random vector we simply get the vector of the expectations And if we take if you want to find variance of a random vector we end up getting a matrix sigma where Σ_{ij} is covariance of X_i X_j for all $i \neq j$ and the diagonal elements are simply the variances of the constituent random variables very good let us move on. Let us move on to the next topic which is a vector differentiation. Let us understand this.

Let us say we have a multi variable real functions, multivariate real functions. So, let us say it is $R^n \rightarrow R$ ok. So, let us say something like f of it is a function of n variables. Now, we know that if it is a function of n variables this denotes the partial derivative of f with respect to x that is fine or with respect to X_j. So, that is the partial derivative of the function with respect to any particular variable X_j, but Now if I tell you that okay I will

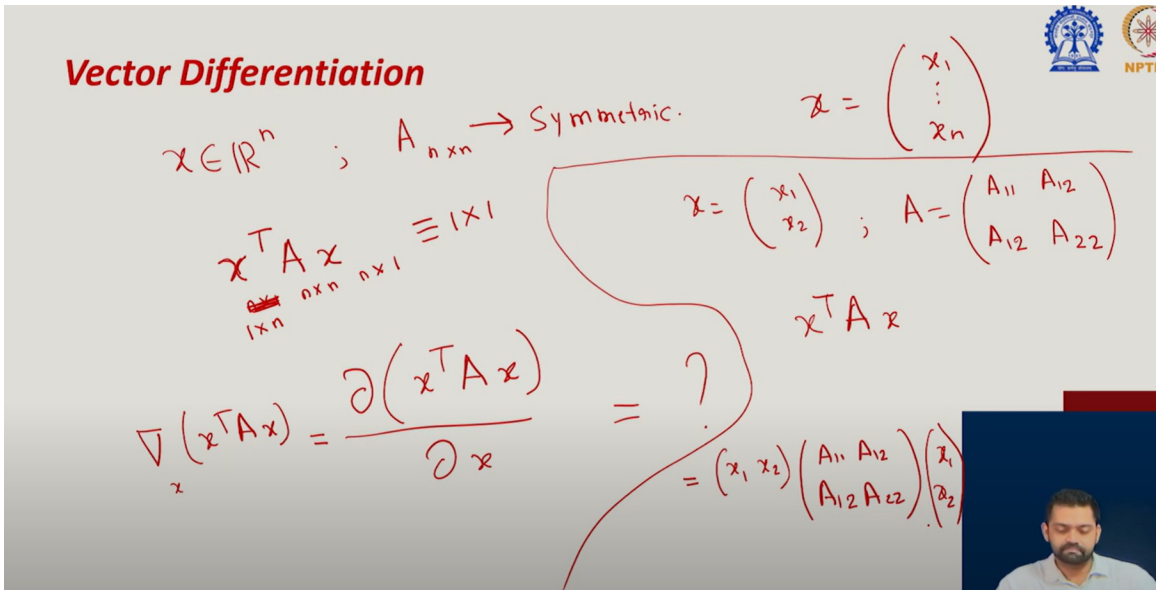


call this the vector x okay this is my entire vector a row vector x.

So now if I tell you what if I have to differentiate this function with respect to a vector x where x is a vector right x is this vector. So what if I want to differentiate with respect to the vector x, then what will you get? Then what you get is another vector which is simply given by this, okay. Well this is often written as the gradient of f with respect to x. so this is the this is this is what the gradient vector is if you have a multivariate function gradient of the function with respect to the input vector is simply given by this very good now let us look at two easy examples and let us remember them because we are going to use them while we study principal components So, let us see first one let us say I have a vector a n dimensional again and I have a vector x again n dimensional ok. So, a is $(a_1 a_2 \ldots a_n)$ x is $(x_1 x_2 \ldots x_n)$.

So, what is a transpose x? so again is what $(a_1 a_2 \ldots a_n)$ and x_1 is this x is this $(x_1 x_2 \ldots x_n)$ so what is a transpose x well that is simply $1, \ldots, n$. Now, if I tell, so this is again, if I consider the a's as the, as constants. So this is again a function of $x_1, x_2, \ldots, x_n$. So this is a multivariable function. Now if I ask, what if I differentiate this function with respect to the vector x? So, what if I differentiate a with respect to x, the vector? what will we have? Let us see not not hard to imagine, we will have $x_1, x_2, \ldots, x_n$ that is what we will have.

Now, what is what is what if I differentiate $\sum a_i x_i$ with respect to x_1, what do we have? So, that is simply a_1 right. So, we have the first element of this vector is this is a_1. What if we differentiate $\sum a_i x_i$ with respect to x_2, what will we have a_2. So, which is nothing but the vector a. So, which means if I differentiate a transpose x with respect to x I end up

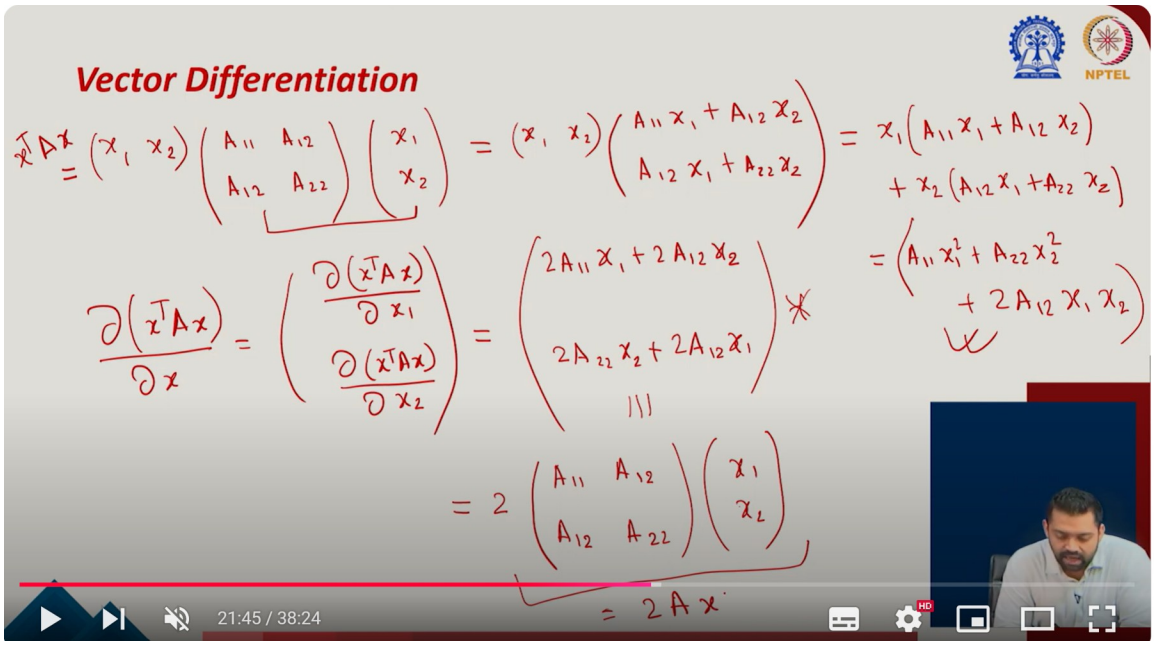


getting the vector a very good.

So, that is that is that is important that is let remember this. Let us look at another form of function Let us say x is a vector again and let us say A is a matrix, A is a n cross n symmetric matrix, okay. If that is the case then what about this function? x transpose A x. Well, this is a scalar function see like we had in the previous case too. What is the dimension of x? x is $n \times 1$, x is this.

So x is $n \times 1$, a is $n \times n$, x is again $n \times 1$, sorry x transpose is going to be $1 \times n$ right, $1 \times n$ and x is $n \times 1$, so that is what we have. So what will be the dimension of this? So this is 1×1, so that is a scalar, so it is a scalar function, good. Now, what if we differentiate this, what if we differentiate this with respect to x or let us say gradient of $x^{tr} A x$, what will this be ok. that is what we want to find out that is our that is that is another problem which we are chasing. So, before we do that let us let us take a simple example instead of n dimensions let us take a little example with 2 dimensions ok and see what we are getting and the same result will actually hold true for n dimensions too.

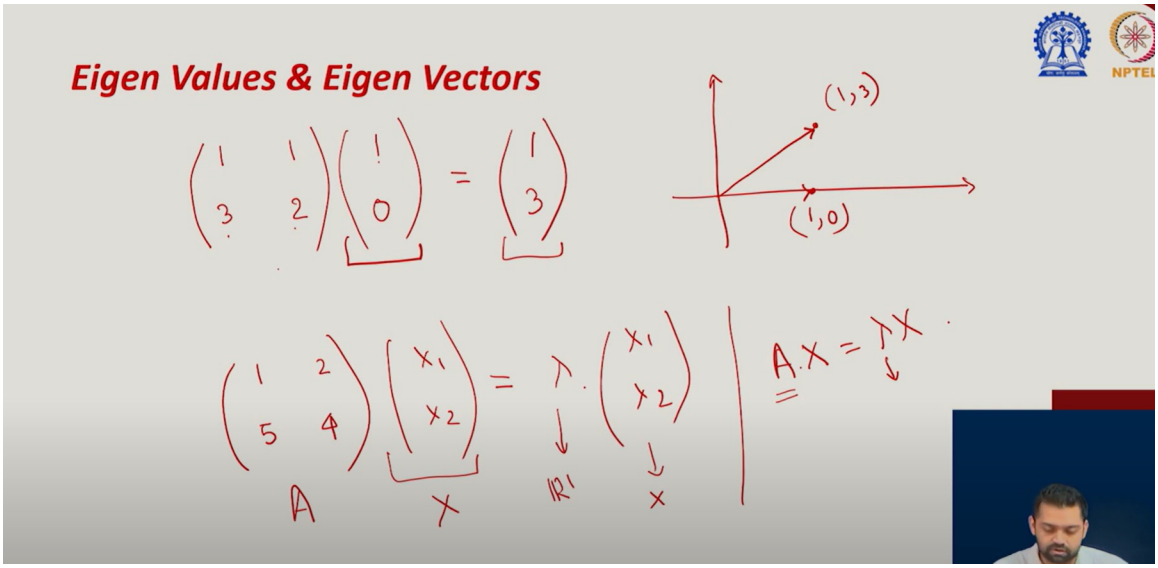
Let us see, let us make it simple, so let us say x is (x_1, x_2), let us say A is, it is a symmetric matrix remember, so $a_{11} a_{12} \ldots$, this and this will be equal, a_{21} and a_{12} is going to be equal right in a symmetric matrix and this is a_{22}. great so if this is the case then what is $x^{tr} A x$ so what is this and then you have (x_1, x_2). So, let us try to multiply this and break it up, let us try to simplify and break this up and let us see what we get, let us try. So, we have (x_1, x_2) and then we had a_{11} If we do this multiplication what will we get? We will have $a_{11} x_1 + a_{12} x_2$ then $a_{12} x_1 + a_{22} x_2$. and then this is a row vector this is a column vector multiply again.



So, you have $a_{11}x_1+a_{12}x_2 + a_{12}x_1+a_{22}x_2$. And if you simplify this you simply end up with $a_{11}x_1^2+a_{22}x_2^2+2a_{12}x_1x_2$ that is what you end up with ok. Now, so this is $x^{tr}Ax$ Now if you differentiate this with respect to the vector x what will you get? So, if you differentiate $x^{tr}Ax$ with respect to the vector x what will you get? You will get with respect to x_1 with respect to x_2 and what is that? If you differentiate with respect to x_1 what will you get? this function if you differentiate with respect to x_1 what will you get? You will simply get $2a_{11}x_1+2a_{12}x_2$. If you differentiate with respect to x_2 what will you get? $2a_{22}x_2+2a_{12}x_1$. Now, this if you take the 2 out this can simply be written as this these two are equivalent you can you can multiply this this you can do this matrix multiplication and see that you will arrive at this matrix.

And what is this? So, this is 2 A x. ok. So, which means if you have if you if you differentiate $x^{tr}Ax$ with respect to x you end up getting 2 A x where A is a symmetric matrix and x is a vector great. So, we have learnt a little bit about vector differentiation we have learnt about random vectors or vectors of random variables and now we have learnt about vector differentiation now our next topic eigenvalues and eigenvectors so I am quickly running through the basics of course you can grab any standard mathematics textbook and read them more in a more detailed fashion Okay, now let us look at eigenvalues and eigenvectors of a matrix. What do we mean by that? Okay, let us begin.

Let us say we have let us say let us take a matrix 1 1 3 2 and let us multiply this with the vector 1 0, okay. So, 1 0 is a vector, this is a vector and I am simply pre multiplying the vector 1 0 with this matrix 1 1 3 2. What will we get? This is 1 into 1 + 1 into 0, so that is 1, then you have $3\times1+2\times0$ that is 3. So what happened? When we multiplied this vector by this matrix, we got another vector. So think of it as an operation on a vector.

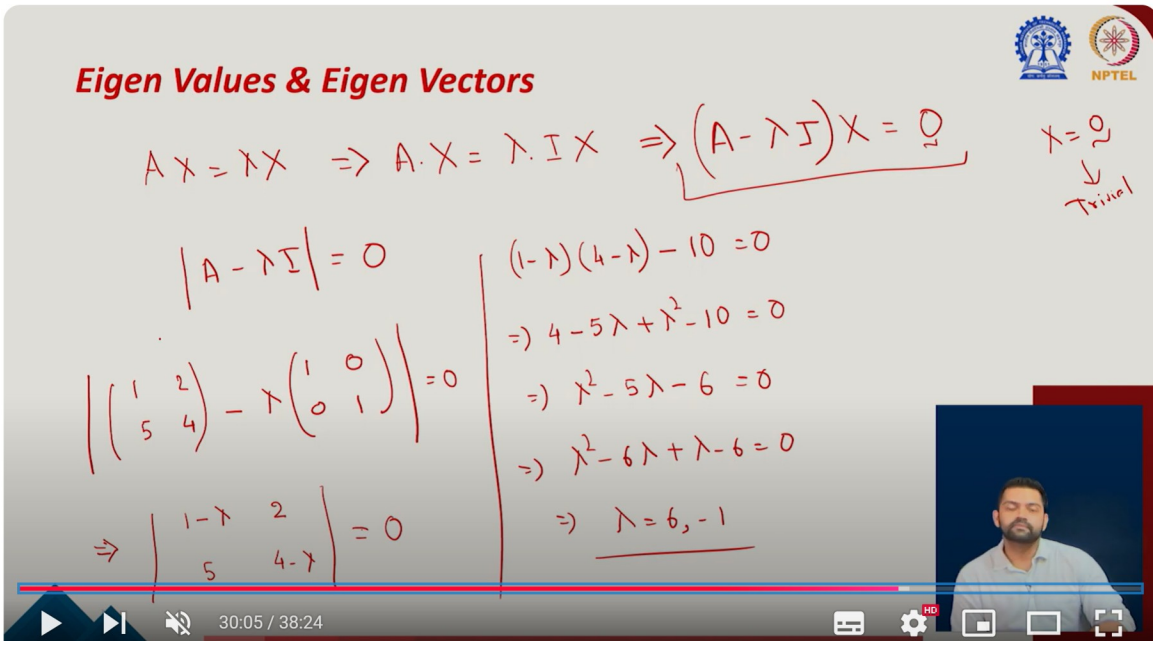


So pre-multiplying by a matrix is actually an operation on a vector, okay. So let's try to understand what actually happened. Let's have a geometrical intuition about it. So in a 2D plane, 1 0 is here this is 1 0 if we pre multiply 1 0 by this particular matrix what did we end up with 1 3 which is this. So, this was the initial vector this is 1 3 and we end up with this vector ok.

So, what happened the vector got rotated not only rotated it also got elongated. what is the length of this vector 1 3 it is it is larger than 1 0 right it is actually the hypotenuse it is root over of 3^2 + 1 square root over of 10 that is the length of the vector 1 3. So, if you if you pre multiply of a particular vector or any vector with a matrix it leads to rotation and change in length ok. fine. Now, what is meant by Eigen vectors of a matrix? Eigen vectors of a matrix are those special family of vectors such that when they are multiplied by any other matrix they do not change directions ok, they do not change direction.

So let us say I have a matrix 1, 2, 5, 4, okay. Let us say I have this matrix and I ask you find the eigenvectors and eigenvalues of this matrix. How will you find it? The answer is the way to find it is you will have to find those vectors which when multiplied by this, let us call this matrix A. so vector like $x_1 x_2$ such that it does not change direction. So, this vector when pre multiplied by a will not change direction.

So, what will it be? It will simply change it may change in length. So, it will simply be a multiple of this vector x. So, it will simply be some λ which is a scalar this is a scalar into is x ok. So, we have if I have a matrix A then a random sorry an eigenvector of A is simply a vector x such that when it is multiplied by A it simply yields another multiple of that particular vector and this multiple λ is called the eigenvalue of A of matrix A.

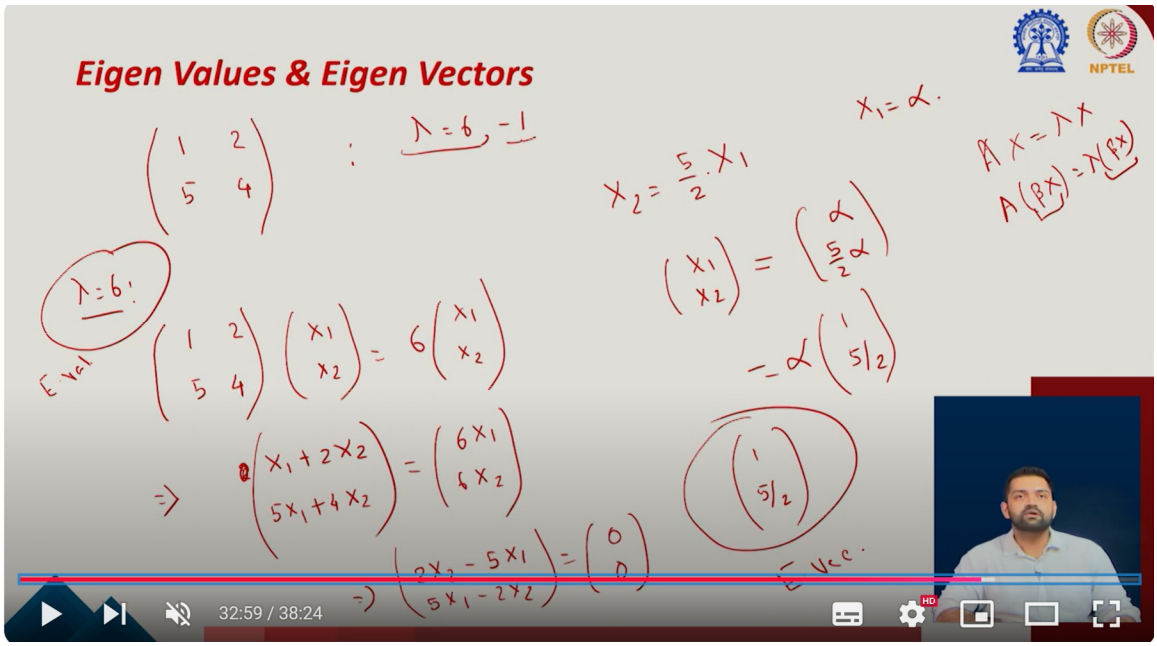


great. So, now that we know if we know matrix A which which is what we do can we find this eigenvalues and eigenvector of a matrix let us see how to do that. So, let us say we have a matrix A and we want $Ax = \lambda x$ we want to solve for λ and x let us see how to do Now this will mean a into x this is λ into the identity matrix into x I can always I can always say that right identity matrix into any vector is that vector. So, that would mean a - λ i into x that is going to be 0 or or a null vector to be precise ok. Of course one solution to this problem is x equal to a null vector, but that is not what we want, we want non-null solutions to this problem, okay. So we are not this x equal to null vector that is trivial, that is trivial, so we are not bothered about that, we are trying to see what other non-trivial solutions are there and when will this equation or set of equations as it will turn out if we write the equations up separately When will they have a non-trivial solution? Well only when the determinant of this coefficient matrix this is 0 did it right ok.

Now the question is what was my a? My a was this 1 2 5 4 remember So what will this determinant look like? So this is 1, 2, 5, 4 - λ what is the identity matrix? Determinant of this is 0. Let us simplify further. So this is 1 - λ 2 5 4 - λ this determinant is 0. Now, let us simplify this further and let us see what we get. So, this is 1 - λ into 4 - λ - 10 that is 0.

So, which means what? So, we are we are landing up with 4 - 5 λ + λ^2 - 10 is 0 or λ square - 5 λ - 6 that is 0. So, what do we have here? So, we have λ equal to 6 and - 1 that is what we have. So, these are the 2 Eigen values of this matrix 1 2 5 4. Now, let us see what are the Eigen vectors. So, 1 2 5 4 has 2 Eigen values what are they 6 and - 1.

Now, let us see. So, that is what we have got. So, let us consider let us start with λ 6 if the eigenvalue is 6 then what will we have right A x is λ x. So, what do we have from



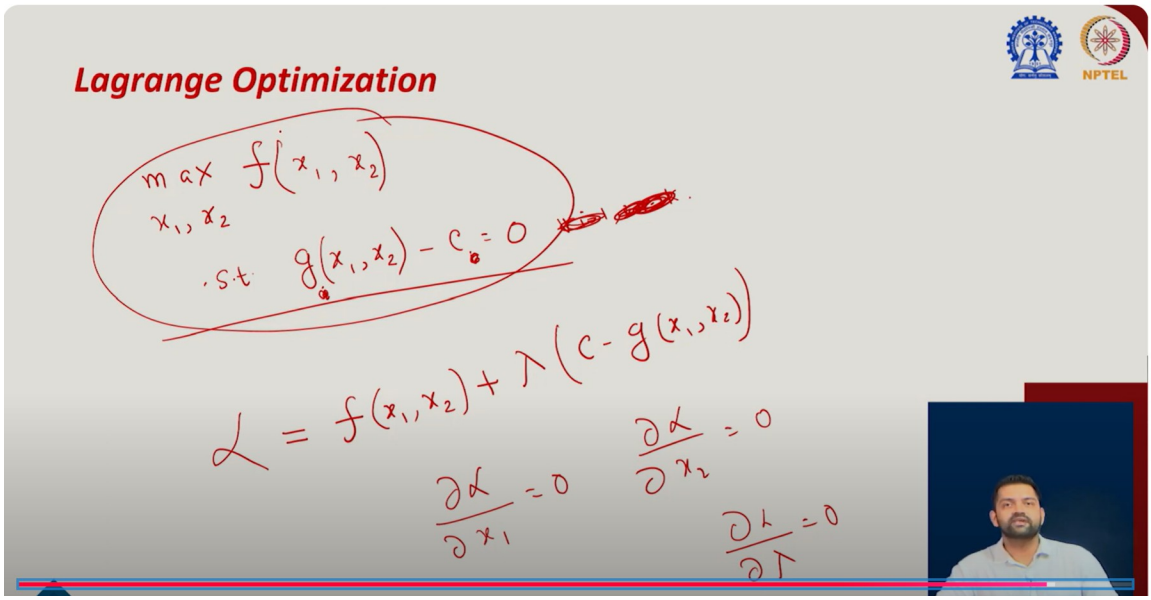
this 2 sorry $x_1 + 2\ x_2$ and this is 5 $x_1 + 4\ x_2$ or we simplify this. So, this is 2 x_2 - 5 x_1 and here you have $5x_1 - 2x_2$ that is 0 0. and what do we get from here.

So, we get x_2 equal to 5 by 2 x_1. So, from this we get x_2 equal to 5 by 2 x_1 ok. So, let us say x_1 is some number α then what is this eigenvector x_1 x it is simply α 5 by 2 α. So, that is so, 1 5 by 2 1 comma 5 by 2 for λ equal to 6 1 comma 5 by 2 is the eigenvector. So, if this is the eigenvalue this is my corresponding eigenvector ok. And when you have an eigenvector any multiple of that eigenvector is also another eigenvector that should be very apparent from this equation right.

If A x is λ x then A into β x will also be λ into β x which means β x will also turn out to be an eigenvector right. So, we basically for every eigenvalue we have an eigenvector and all those and all multiples of that eigenvector are they are also eigenvectors So, similarly for λ equal to - 1, I can also compute my family of eigenvectors, great. So, we have learnt about eigenvalues and eigenvectors. Now we will talk about our fourth topic of our map prerequisites. So, first we have learnt about ah random ah vectors or vectors of random variables, then we learnt about vector differentiation, then we learnt about eigenvalues and eigenvectors and now Lagrange optimization.

So, what is Lagrange optimization? So, suppose we have a function f So, Lagrange optimization is a way of constrained optimization. So, I will just give you an example. Let us say I have a function f x 1 x 2 and I want to maximize this function such that there is a constraint. like this or there are multiple constraints.

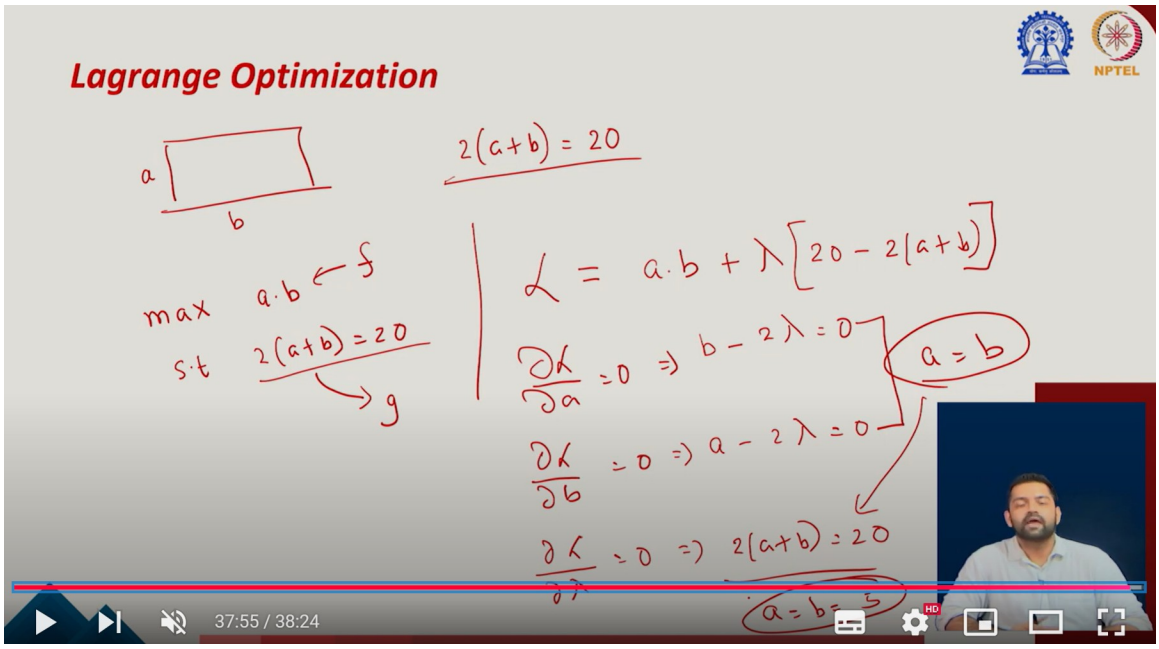
So, let us saygi or the ci. So, let us say there are k constraints or something like this or let



us say two constraints. So, in this case we will just consider one constraint let us let us keep it simple. Then what do we do? We define a Lagrangian So we define a Lagrangian, what is that? It is this, this is called the objective function, f is the objective function and then we simply write this as I am not getting into the mathematical nuances of the technique, I am just giving you the workable and just passing on a workable skill here. So, we just write the Lagrangian in this manner and then we find the partial derivatives. and then we solve these equations and whatever we get is the solution to this optimization problem ok.

So, let us see let us look at a very simple problem let us say I tell you that consider all rectangles which have a perimeter of 20 which amongst those rectangles have the highest area very simple problem. So, so let us say there is a rectangle a b length is b the width is a then what is the perimeter 2 a + b that is 20 that is Now the question is what is the maximum area of such a rectangle? So, what is my problem then? My problem is maximize the area which is a into b such that the perimeter 2 a + b is 20. So, this is my f the objective function, this is my g the constraint ok. so what will we do of course this can be solved in in just a line, but I am just writing down to illustrate what are like how a Lagrangian works. So, I write down my objective function + the Lagrangian this λ is called the Lagrange multiplier by the way λ 20 - 2 a + b ok.

Then we do what we do so if you differentiate with respect to a what will you get you will get b - 2 λ is 0 if you differentiate with respect to sorry with respect to a you got this with respect to b what will you get you will get a - 2 λ is 0 and from this what will you get you will get 2 a + b is equal to 20 you will get back the constraint which you had the g What will the first two tell you? They will tell you that A is equal to B. Now you substitute this into this equation and you will get A equal to B equal to 5 and that is your



solution, okay. That in a nutshell is Lagrange optimization. All these techniques, the four topics which I have touched upon in the last 30-35 minutes Of course you should ideally grab a proper mathematics textbook and learn it properly, but I tried to cover it as much as possible as a prerequisite of what follows, which is the principal component analysis technique.

That is what we look into in the next lecture. Thank you. See you in the next lecture.