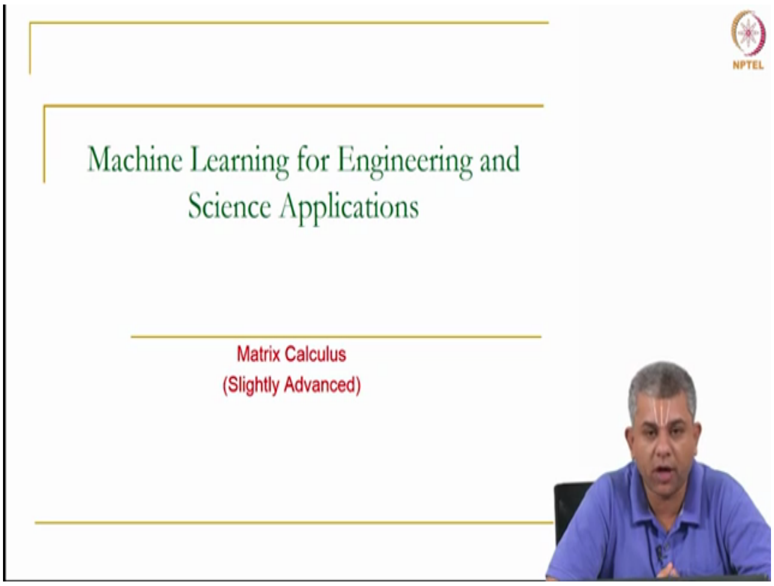
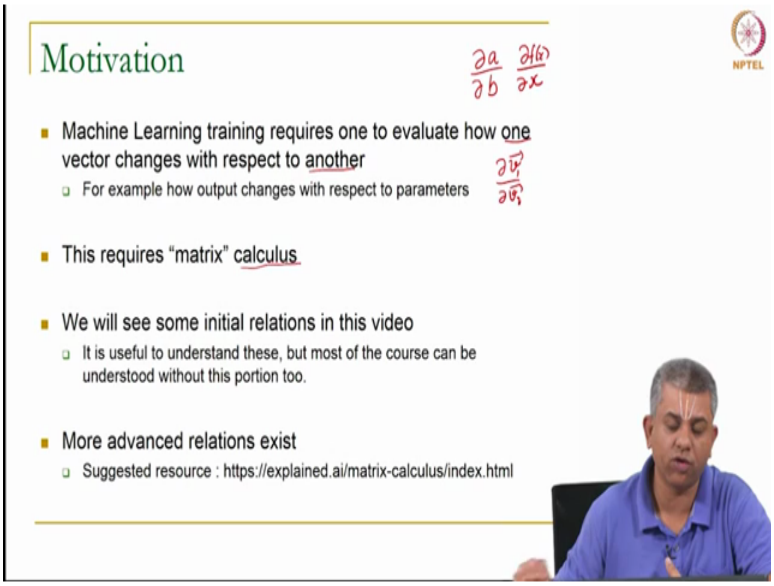
Machine Learning for Engineering and Science Applications
Professor Dr. Balaji Srinivasan
Department of Mechanical Engineering
Indian Institute of Technology, Madras
Matrix Calculus (Slightly Advanced)

(Refer Slide Time: 0:15)



In this video we will be looking at matrix calculus, this is a short video the portions are slightly advanced, once again like with some portions of the probability series you are free to skip this I would still recommend that you go through this and see some of the relations if you are not able to understand them or fully exploit them that is fine because we will be not using this for most part of the course about 90 percent of the course can be done even without understanding this very well.

(Refer Slide Time: 0:46)



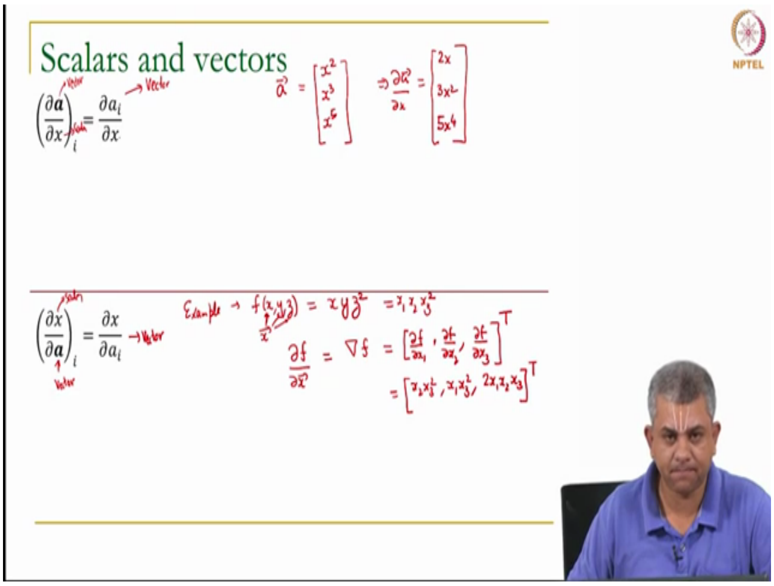
So here is motivation for why we are looking at it, please remember that as we had said in the first couple of weeks machine learning basically requires you to take some input vector and change it into some output vector. Now what will often happen during training is you will notice that the input vector that you have given does not quite give whatever required output vector you have, whatever you would require.

So for example you know if your output changes with respect to some set of parameters that you have, you would like to know how much will the output change provided I turn a few knobs or I change a few parameters. So in such cases you basically need to know how one vector changes with respect to another vector or another some other parameter. So a standard way of how we measure, how one quantity changes with respect to another of course is the partial derivative if you have two scalars you know very well how to change you know or find out how one function changes with respect to a parameter x. So all this is of course a subset of calculus.

Now we are going to slightly extend this idea into matrix calculus which basically means how does one vector change with respect to another vector and how do you parameterize this and what are some of the basic relations this is going to be a very initial or preliminary class I am going to use only some of the relations that we will require in terms of machine learning, of course much more advanced material exists as I said before in the introduction it is useful to understand these but in case you do want to go ahead with the course even without understanding this material that is fine you will be able to extract 90 percent of the information of this course anyway.

Of course in comparison to the relations that I am showing there are many more advanced relations which do exist. A good source is the one that I have flashed on the screen right now and there are sources within this website also which go into greater detail.

(Refer Slide Time: 3:00)



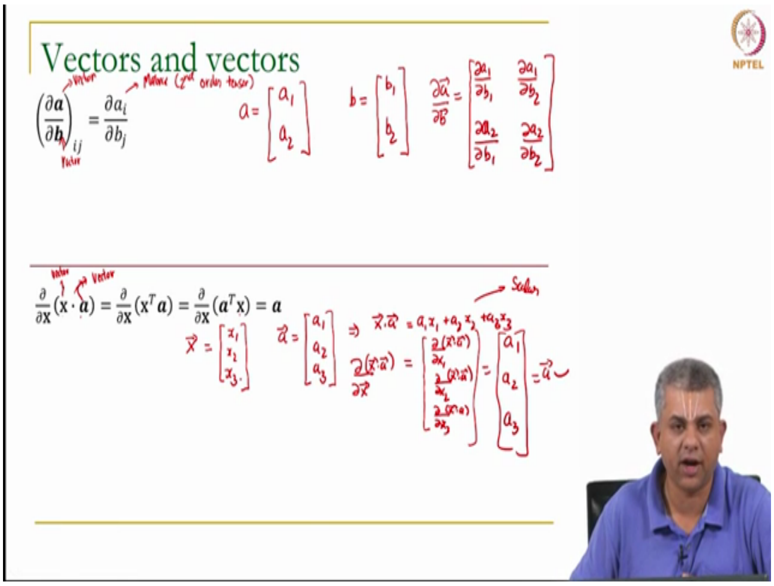
So let us look at one simple case, we will assume of course that you know how to differentiate one scalar with respect to another, but let us say you have a scalar with respect to a vector or vector with respect to a scalar in this case, this is del a vector with respect to x remember this is a vector and this is a scalar. So if you differentiate a vector with respect to a scalar the result that you get is a vector.

So for example let us say a vector is x square x cube x cube r 5 and x is a scalar, so del a vector del x also has three components which is going to be 2x 3x square and 5x to the power 4 which is what is represented here the ith component of this vector is simply del ai del x. So I hope this portion is clear, you could have the reverse case where you have a scalar differentiated with respect to a vector, so del x del ai an example would be something like.

So this is a scalar function, it takes three inputs x, y, z we could call this x vector or if you are interested we can even call this x 1, x 2, x 3 square and then del f with respect to del x vector this is of course what we call the gradient, this is now a vector. The first component is going to be del f del x 1, the second component is going to be del f del x 2, the third component is going to be del f del x 3, I will put a transpose here because this is right now a row vector you can turn it into a column vector.

So del f del x 1 is x 2 x 3 square, x 1 x 3 square, 2x 1 x 2 x 3. So this is differentiation of a scalar with respect to a vector. So a gradient is a prototypical example of some such thing it also results in a vector.

(Refer Slide Time: 6:06)



Now let us look at a slightly more involved case the differentiation of one vector with respect to another vector, where does it occur physically? I mean of course in machine learning we might not necessarily look at you know physical examples but just for physical intuition let us say you have velocity in a particular room of the air and you want to differentiate it with respect to the position. So in each point at each point you can have x velocity, y velocity, z velocity changing with respect to the location x, y, z so as I move to a different point all three components will change.

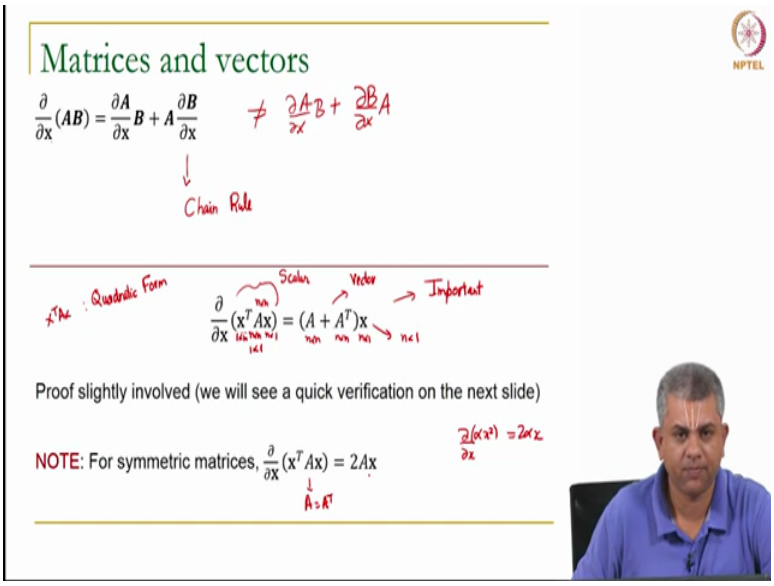
So this is a vector differentiated with respect to another vector and what it results in is a matrix or a what we call a second order tensor. It is actually a fairly relationship ith component of i differentiated with respect to jth component of b. So for example if we have a is the vector a 1 a 2 and b is the vector b 1 b 2 then del a vector by del b vector and del a 2 with respect to b 2, so this is a matrix and that is what is written here del a del b ij is equal to del a i del b j.

This relation we will actually be using you know a little bit more. So this is differentiation of a dot product with respect to x of course this is actually speaking, so let us say this is a vector and this is a vector the product is of course going to be a scalar. So this is a special case of the previous example we had seen, but unlike the previous case an x is actually occurring here, okay.

So remember x dot a can be written as x transpose a where x is a vector you transpose the matrix a row matrix multiplied by a column matrix, it can also be written as a transpose x both are the same because the product is actually a scalar and you can show that del del x of x dot a is equal to a vector. So I will just quickly show it to you in a special case you can also show this in general.

So let us say x vector is x 1 x 2 x 3 and a vector is a 1 a 2 a 3, so this means x dot a as you know is a 1 x 1 plus a 2 x 2 plus a 3 x 3 remember this is a sum and now you have a scalar. Now del del x vector of x dot a is going to be del del x 1 of x dot a as we saw in the previous slide del del x 2 x dot a and del del x 3 x dot a. Now you can immediately see del del x 1 of this is a 1, del del x 2 of this is a 2, so this is a vector, okay. So hence you can proof this I have shown this in the 3 dimensional case you can of course show this in n dimensions quite easily.

(Refer Slide Time: 10:05)



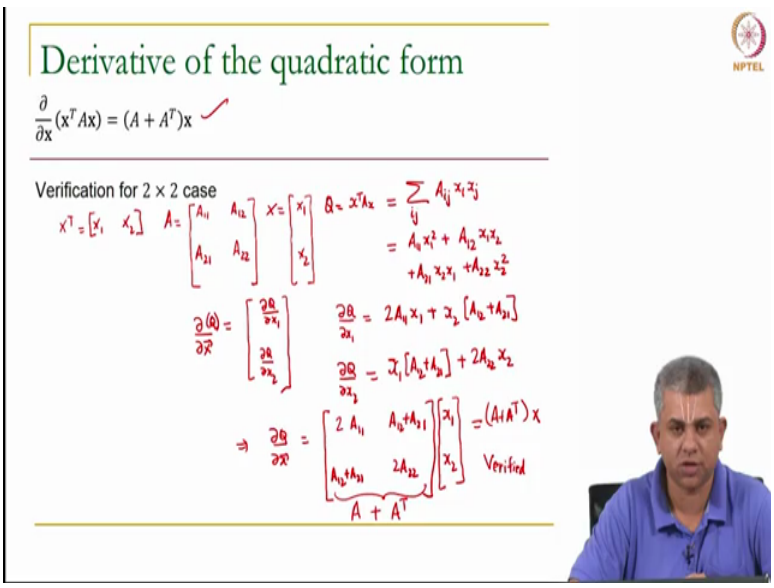
Now let us come to the next level of evaluation which is matrices with respect to vectors. So suppose you have product of two matrices AB and you are differentiating it with respect to a vector x, it could be a vector or a scalar in either case the relationship that is shown here actually holds, del a del x times b so this is the equivalent of chain rule remember that the order AB never changes. So this is del a del x times b, do not write this as you know del a del x times b plus del b del x times a so this does not work because matrix product in general does not commute. So the chain rule you have to be careful about the order of the matrices being used.

And this the relation I am showing now the result I am showing now is important along with the dot product rule that I showed a little bit below. Remember that let us say A is an n cross n matrix then x transpose a n is actually a scalar, x transpose Ax is called a the quadratic or a quadratic form you can review your linear algebra lectures where this was discussed so x transpose Ax is a quadratic form this is all put together is scalar, why? Because this is n cross n, this is n cross 1 and this is 1 cross n. So if you multiply all of them you will get 1 cross 1 so you will see natural places where this tends to occur.

Now what happens is the derivative of this with respect to x remember derivative of scalar with respect to a vector also has to be a vector and indeed it is a vector so this is n cross n, n cross n, n cross 1 so the result actually is a n cross 1 vector, scalar with respect to vector is actually a vector. So the proof of this we are not going to do this, so the proof is actually slightly involved I will show you a quick verification just like with the dot product case on the next slide.

A useful relationship is that in case you have a matrix which is symmetric. So symmetric A means A is equal to A transpose, so this relation simply becomes 2Ax and it looks remarkably like our scalar formula which is d by dx of let us say some alpha x square is equal to 2 alpha x, of course this is for scalar x here you have to be little bit careful, this is x transpose Ax and you get 2 times Ax for symmetric A.

(Refer Slide Time: 13:11)



So let us look at the quadratic form and look at a quick verification, so we will look at the verification for a 2 by 2 case. So let us assume A is of the form A 11 A 12 A 21 A 22 of course x transpose will be x 1 x 2 and x can be written as x 1 x 2. We call that x transpose Ax is essentially summation over i and j of all products of the form A ij x i x j, you can of course multiply it out using the usual matrix multiplication and find this, but if you do this so you will get A 11 x 1 square plus A 12 x 1 x 2 plus A 21 x 2 x 1 plus A 22 x 2 square.

So now if I have to find out del del x vector of this so let us call this Q, so as before this is del Q del x 1 del Q del x 2. So we know del Q del x 1 is equal to 2A 11 x 1 plus x 2 times A 12 plus A 21. Similarly del Q del x 2 is equal to this term does not contribute anything you get x 1 times A 12 plus A 21 plus 2A 22 x 2. So now this can be written as please notice this 2A 11 A 12 plus A 21 again A 12 plus A 21 2A 22 times x 1 x 2, this matrix of course is A plus A transpose, so this you can quickly see if you do A transpose it is going to be A 12 A 21 you add the two you are going to get this.

So this is A plus A transpose times x, so you have verified these relationships. So we will be using these relationships OFF and ON during the rest of the course, thank you.