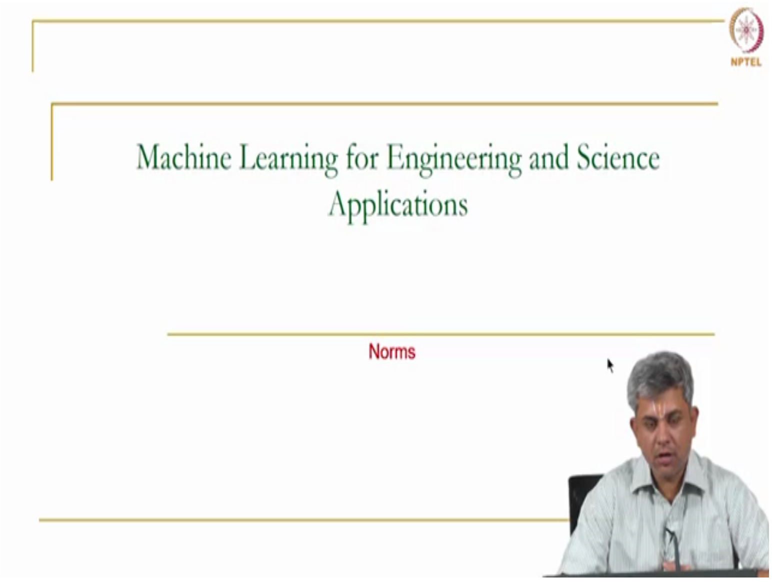
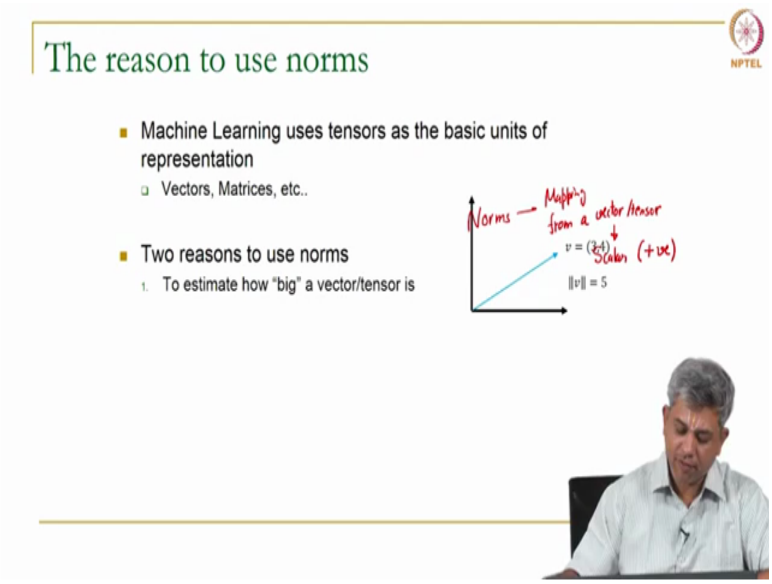
Machine Learning for Engineering and Science Applications
Professor Dr. Balaji Srinivasan
Department of Mechanical Engineering
Indian Institute of Technology, Madras
Norms

(Refer Slide Time: 0:15)



In this video we will be looking at an important idea, this is the idea of norms this is one idea that we will be using throughout the rest of this course.

(Refer Slide Time: 0:28)

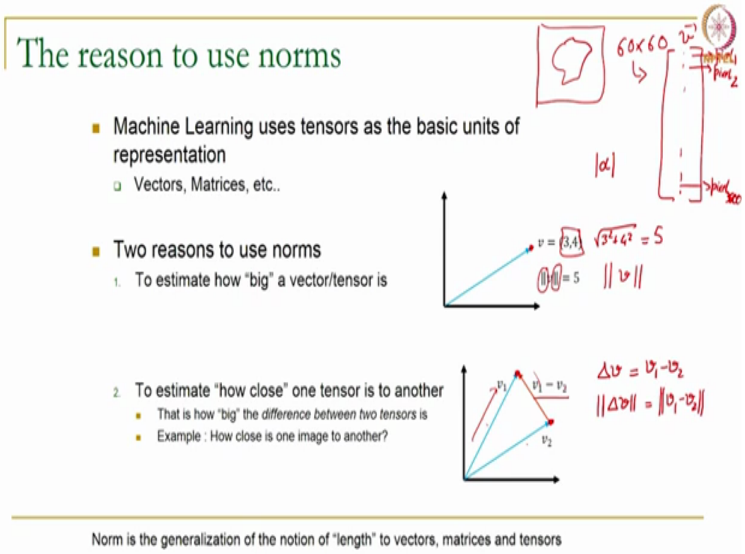


So norms are an idea in linear algebra or in general whenever we deal with tensorial quantities. The basic reason why machine learning and many other fields use norms is because we usually use vectors or matrices as our basic units of representation. As we saw in the last video we tend to use vectors and matrices very very often basically because that is what we use in order to measure or in order to represent images, sounds or anything in fact anything that goes our input or output is usually measured by vectors and matrices.

So there are two basic reasons that we use norms, one is to find out how big or small a particular vector or tensor is sometimes we need to estimate the size of something. Now for a scalar or if it is a scalar like a weight or pressure or temperature there is one single number by which we can get the idea of how big this thing is, whether it is negative or positive the absolute value usually denotes what the size is for a scalar.

For a vector we have no such single number of course vector is a bunch of numbers but suppose you need a single number. So norms sometimes can be thought of as a mapping from a vector or a tensor to a single number to scalar and actually this is a positive scalar. So we will see how to do that in the rest of this video, there is another reason for which we use norms.

(Refer Slide Time: 2:18)



So for example let us say you have a vector of this sort, usually we will denote the size or the length of this vector as square root of 3 square plus 4 square this is 5. So the usual notion of length a norm is denoted by this sign usually a double bar sign just like for scalar we use single bars for absolute values, for norms we tend to use this double bar, some people use single bar also so we will see this notation a little bit later on in the video.

So whenever you hear me say norms please think of you know a simple vector for which you are trying to find out the length essentially you are trying to find out one single number that will represent the size or how big a particular vector is, there is another reason for which we use norms which is to try and estimate how close one vector or tensor is to another, okay. So once again I would like you to think about the idea of images in order to show something which is qualitative where you can estimate this.

So please remember if you recall what we did in the previous videos, we had looked at a whole image. So let us say you have an image of a cat or something and this is a 60 cross 60 image, we saw that this can be unrolled into a single vector which is of size 3600, each of these represents one pixel, okay. So you have 3600 pixels, so it can be written as a vector of dimension 3600.

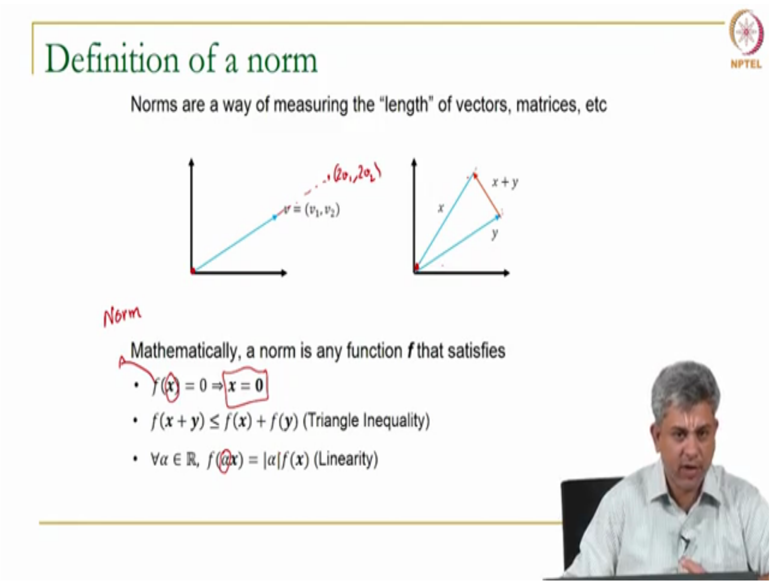
So now you cannot really imagine this but let us assume that instead of this (this would) this is just two numbers so it is as if it is an image of just two pixels, but suppose you have a one whole image of 3600 pixels now you can start thinking about you can now imagine this is one image and this is another image, okay of course we are representing it in two dimensional space, so each of these points is a vector which represents one image and suppose you want to find out is this image close to the other image, okay now how would you do that?

So that idea also basically would be how big the difference between these two vectors is we know of course that the difference between two vectors is another vector. So if you have this vector v 1, this vector v 2, v 1 minus v 2 is another vector and I could find out delta v is v 1 minus v 2 if I find out the norm of delta v, okay or the length of this vector which is the difference of these two vectors that will tell me how close the two images are.

So a norm is supposed to represent both these ideas or atleast its used when both these ideas which is essentially if you can somehow define one single number to represent the size of one whole vector or one whole tensor then you have the idea of norm. So usually like I said just now you can try to find out how close one sound is to another if you have two representations, how close one word is to another, how close one image is to another provide you all of this can be represented as vectors and you can find out the norm of the difference between the two vectors, okay.

So now let us see how to go about doing this. The norm is actually a generalization as you can probably figure out of the notion of length, the idea that we have of length for simple scalars can now or size of simple scalars to vectors, matrices and tensors.

(Refer Slide Time: 6:10)



So let us say you have a vector all my example which I show on the slide will be in 2D of course you can imagine this being extended to multiple images, okay. So the numerical example I will be taking would be that of a 3D vector. (Mathematics) Mathematically what we will be doing is we will be trying to generalize we will find out what the specific properties of length are which makes it intuitive and the useful notion for us in real life.

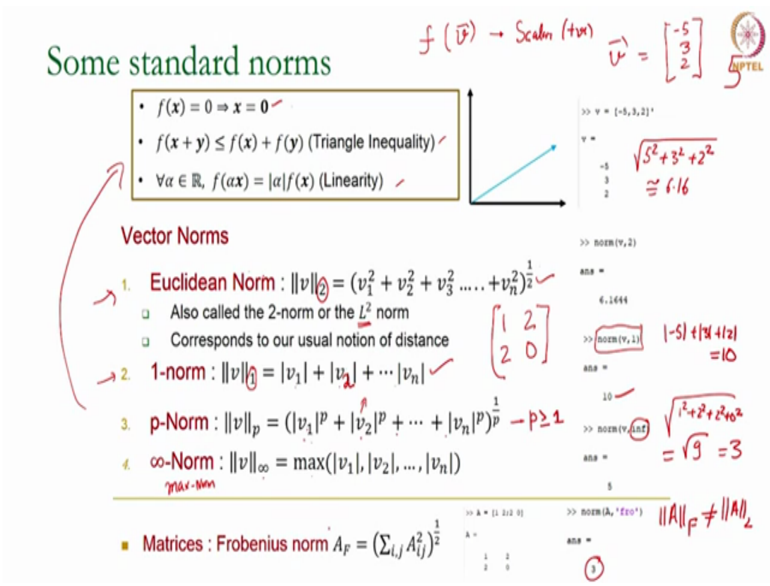
So the first notion which is very important is if you have a vector whose length is 0, then that means it is a 0 vector. So the only vector which is of length 0 is essentially this vector which is right at the origin, okay so that is the first property that any norm should satisfy that is if the vector has length 0, then it must be the 0 vector, okay. So this is the definition of norm that we will be using here.

The second property is the property of the triangle inequalities, so let us say you have two vectors please notice I have flipped the arrow here just in order to be consistent with the mathematics that I will be using. So let us say the first vector is x and the second vector is y, okay. Now we know that x plus y has to be this vector here, okay going from here to here it is a simple vector addition rules.

Now what the triangle inequality rule for the norm says is that the length of this has to be always less that the length of this plus the length of this, we know this from the normal triangle inequality that we use for triangles right from schools, the length of two sides is always going to be larger than the length of the third side, okay the sum of two sides is always going to be larger than the third side that is because the shortest distance between any two points is a straight line.

So if I want to go from here to here, you know if I go that way that will always be longer than this, so this is the normal triangle inequality rule it is represented as f, f you can think of a sub function which represents a norm, norm of the sum of two vectors is going to be less than equal to the norm of the individual vectors, okay it is a very very important property. The third property that a norm satisfies is that of linearity, what it means is if I take a vector and simply scale it up, take a string extend it by two times each of the coordinates will increase by a factor of 2, so let us say if I increase it by a factor of alpha then its length also increases by a factor of alpha, these are the three properties that any norm satisfies.

(Refer Slide Time: 9:00)



Now based on these three properties that we just saw here the idea of 0, the idea of triangle inequality and the idea of linearity, what we can do is we can derive many many many different functions that satisfy this, okay. So remember f the norm takes in a vector gives a scalar which is positive and you need to define you can define many functions which satisfy these three properties, okay.

So let us take a simple example, so we are taking the example of a vector which is minus 5 3 2, so let us say we have a 3 dimensional vector and we will see various norms that can be used for this simple vector. The first and the most obvious norm is called the Euclidean norm sometimes the Pythagorean norm, the Euclidean norm you will notice has a subscript 2 the reason for the subscript will become obvious very shortly, okay.

So you have a vector all it is root of the sum of squares, so you take in this case you would do square root of 5 square plus 3 square plus 2 square essentially what we usually call the length of the vector, okay this is also called the 2-norm or sometimes also called the L 2 norm, okay the reason for the L will not go over but usually you will see this term being used a lot of times 2-norm or L 2 norm, okay.

So what is the L 2 norm of this case? It usually corresponds to our notion of distance so you can immediately find out that this is equal to approximately 6.16. A similar norm is called the 1-norm please notice the subscript here, all it is instead of squaring and taking square root you simply add the absolute values, okay. So in this case our 1-norm would be very obviously I have written a MATLAB command here, but you can do it by hand in this case all it is absolute of minus 5 plus absolute of 3 plus absolute of 2 which is equal to 10.

Now using these two you can generalize to the idea of what is called a p-norm, p-norm is simply absolute of v 1 to the power p plus absolute of v 2 to the power p remember all these are components, okay the whole thing to the power 1 over p, okay. So you will notice that this covers both the 1-norm and the 2-norm and this kind of definition is valid for p greater than equal to 1.

So usually you cannot define let us say a half norm or something but 1 and so on and so forth you can define all other norms. As it turns out these two are extremely useful norms, there is also a third norm which is very useful which is called the infinity-norm or sometimes called the max-norm, so the max-norm simply is find out the maximum component in absolute values, so in our case max of absolute of minus 5, 3 and 2 which should basically the infinity-norm will simply be 5.

So you can check that MATLAB has a command norm v comma inf, inf gives you a maximum of 5. Now what is interesting is you can actually see the max-norm as a limit of the p-norm as you keep on increasing p, okay. As you keep on increasing p, let us say the v 2th component was the largest component what will happen is all the other terms will become very very small as you keep on increasing the power in comparison to (v 1 to the power p) v 2 to the power p will be very very large as p becomes large and in the limit of infinity this is the only term that survives and once you take a 1 over p what survives is the maximum-norm. So this is either called the infinity-norm or the maximum-norm.

Now I want to emphasize that the most natural norm atleast the one that we think of very naturally is the 2-norm, none the less 1-norm or infinity-norm can also be useful. Please notice that each of these norms or all of these norms satisfy these three properties, okay we are not going to prove this, we know that the Euclidean norm satisfy this by intuition just as a quick check for example you can check that if you take the infinity-norm it is definitely going to satisfy this, the only way in which can infinity-norm can be 0, that is the maximum of the absolute value of something can be 0, if all the components were exactly 0.

Similarly if the sum of absolute values is equal to 0, the only way that is possible is each of this individual this should be v 2 (I am sorry) each of this individual components is 0, okay. So these three properties are satisfied by all of these three norms. Now all these norms as I have showed them apply to normal vectors you can actually extend this idea to matrices also, the idea of norm is true for vectors, tensors and matrices. The definition remains the same or atleast the properties remain the same x instead of being a vector becomes a matrix.

You also have 1-norm, 2-norm, infinity-norm for a matrix, but in machine learning the most common norm that we use is what is called the Frobenius norm, Frobenius norm is very similar to the Euclidean norm all it is you take all the components of a matrix, so let us say I have a matrix here 1 2 2 0, the Frobenius norm of the matrix is square root of 1 square plus 2 square plus 2 square plus 0 square basically some of the squares take the square root, okay that is the Frobenius norm, in this case this is square root of 9 which is equal to 3, okay. So that is the Frobenius norm.

Please notice the Frobenius norm denoted by A subscript F is not the same as the matrix 2-norm, okay there is some such thing as the matrix 2-norm or the matrix you know L 2 norm that is not the same as the Euclidean norm, so there is a slight difference there none the less the Frobenius norm is probably once again the most common thing that you will think of, immediately if you want to find out one number that represents the size of the matrix.

So this is the idea of the norm we will be using this repeatedly again and again through the rest of the course, one of the main uses that we will be using it for is you know as you are using iterative procedure for a vector, okay so suppose you are trying to find out some particular parameter vector or some particular image and you are trying to go slowly go there through an iterative process your initial guess is bad and you are slowly getting there, you want to find out how close each guess is to the final guess and one of those ways to find out is as we saw earlier find out the difference between the two and take there norm. So we will be using this repeatedly through the rest of the course, thank you.