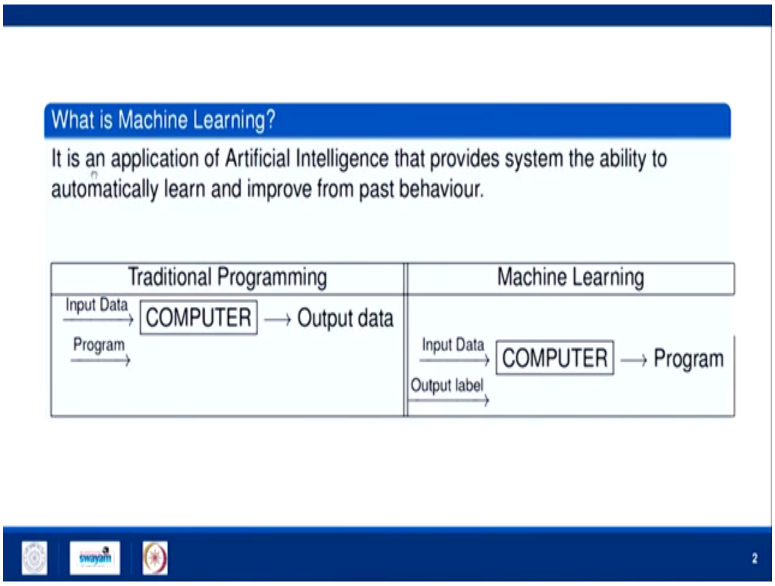
Essential Mathematics for Machine Learning
Prof. Sanjeev Kumar
Department of Mathematics
Indian Institute of Technology, Roorkee

Lecture – 01
Introduction to Course and Vectors

Hello friends. So, I welcome you all to this course on Essential Mathematics for Machine Learning. So, in this course, we will tell you some mathematical concepts those are really helpful or those are really important in the area of machine learning, deep learning, artificial intelligence and so on. So, I will start with the basic definition of machine learning.

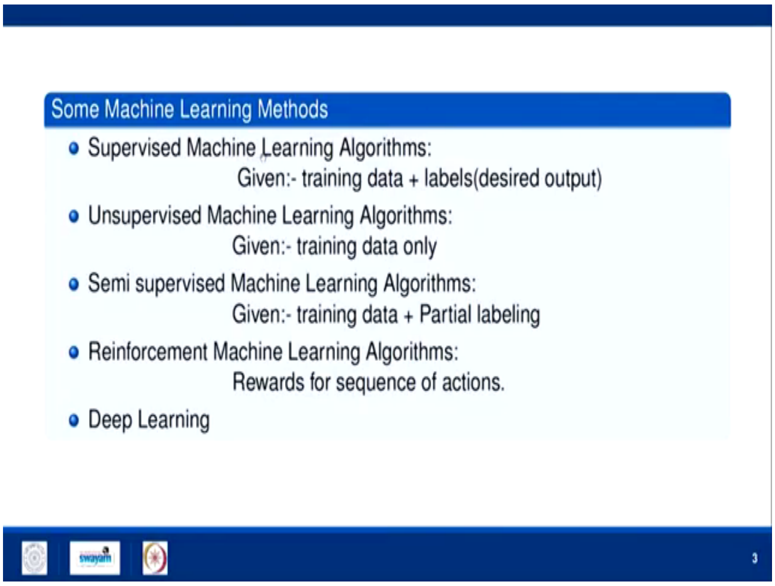
(Refer Slide Time: 00:55)



So, machine learning is basically an application of artificial intelligence. So, in short, I will say ML for Machine Learning. So, ML is an application of artificial intelligence that provides system to the ability to automatically learn and improve from past behaviour. So, we learn from our past behaviour and we take our decision in future or in current time. Similarly, we want to give similar ability to machines, so that based on the past record, they can make some decision in present or in future.

All of you are aware with traditional programming. So, what we use to do in classical programming? We are having input data and program. We feed these two things to computer. And what we get? Output data. However, machine learning is different. In this, what we are having? Input data and output, we feed both of them to the computer, and we will try to model our program ok. So, whatever happens in past using those past records, we will try to model the phenomenon from which we got output for a given specific input, so that is basically machine learning.

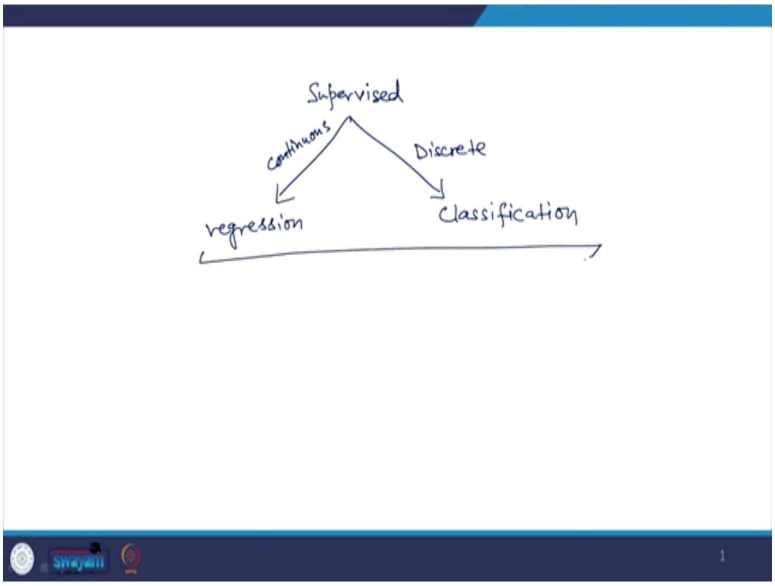
(Refer Slide Time: 02:30)



There are different types of learning machine learning methods. Now, first methods are supervised machine learning algorithms. So, in supervised machine learning algorithms, we are having input as well as output levels for some of the training data ok. For some of the data and that data is called training data.

So, based on the input values and output levels, we will try to learn using the training data; and for testing or for the data for which we do not know the output using that training or learning we want to predict their output level ok. So, this kind of supervised learning is called this kind of machine learning is called supervised machine learning.

(Refer Slide Time: 03:21)



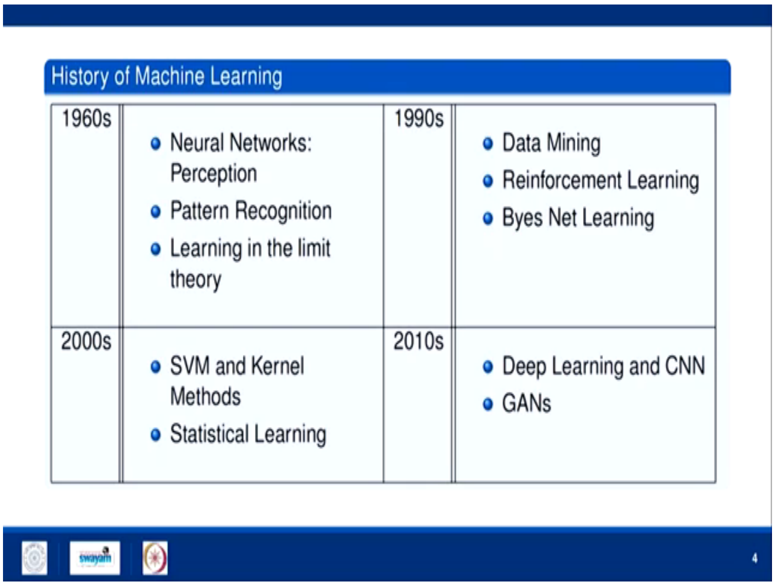
So, again it is of two types. So, supervised, so we are having two types of thing in this; one is apt is on continuous we are learning a continuous function; then it is called regression. And if we are learning a discrete function then it is called classification.

The second types of methods are unsupervised machine learning. So, in unsupervised machine learning, we are having only training data means only input we do not know about the output levels. Now, from those training data, we will try to expect some of the properties. And based on the on the similar properties of few of the data, we want to group them together. We want to make a cluster of them. And such type of clustering is called unsupervised machine learning algorithm.

Then we are having semi supervised machine learning algorithms. So, in this we are having training data input as well as partial labeling. Note the complete labeling as we are having in supervised learning for the training data. So, based on those partial labeling, we will try to find out level for the rest of the training data pattern, and then we include them our training data.

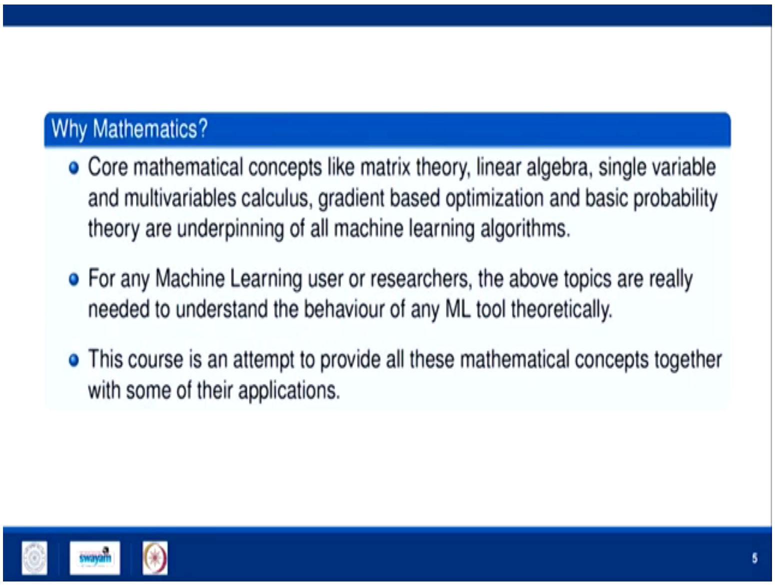
Then we are having reinforcement and machine learning algorithms. So, this rewards for the sequence of actions in this kind of thing. And nowadays very popular deep learning. So, in deep learning, we are having supervised as well as unsupervised learning algorithms. And there as well as reinforcement machine learning algorithms. And there we use the deep architectures of the tools like neural networks, etcetera. So, we are having multiple layers, multiple hidden layers. And we are having high-end processors and using AGPU; using those we try to out ah regression or classification.

(Refer Slide Time: 05:42)



So, if you talk about the history of ML, it is in 1960s we are having perceptron model, pattern recognition, learning in the limit theory. Then in 90s, we were having data mining, reinforcement learning, byes net learning. In 2000s, we are having SVM, kernel method, statistical learning. And from 2008-09 means in nowadays we are having deep learning and convolution neural networks. And then we are having GANs type of thing, GANs.

(Refer Slide Time: 06:20)

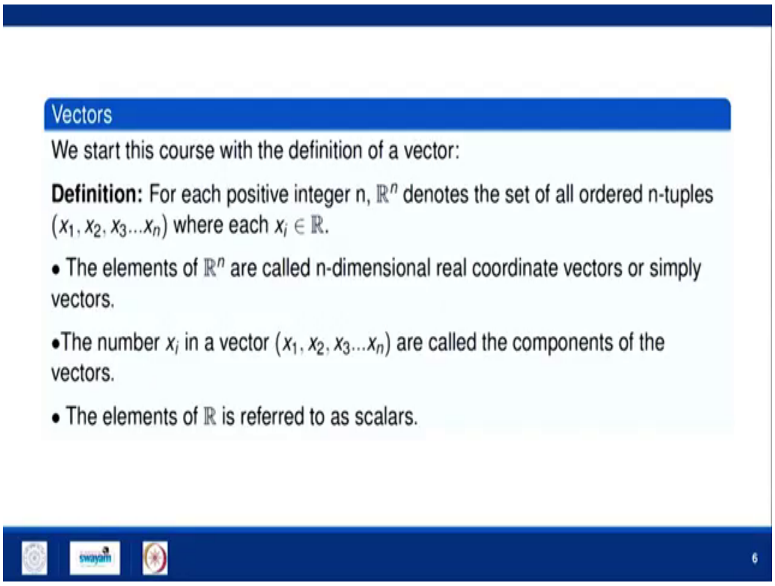


Now, this course is essential mathematics for machine learning. So, why mathematics is required in machine learning? So, core mathematical concepts like matrix theory, linear algebra, gradient calculus, optimization, probability theory are underpinning of all machine learning algorithms, each and every machine learning algorithms are having some concept coming from these topics which I told you, and that is why in this course we have include all these topics. Those are directly relevant to your machine learning algorithms.

So, we will start this course with matrix theory, linear algebra, and then we will go to the algorithm based on the eigenvalue and eigenvectors of linear transformations like principal component analysis, linear discriminate analysis and so on. So, in first 20 lectures which I will take we will cover the linear algebra and matrix theory, and relevant topic from the machine learning.

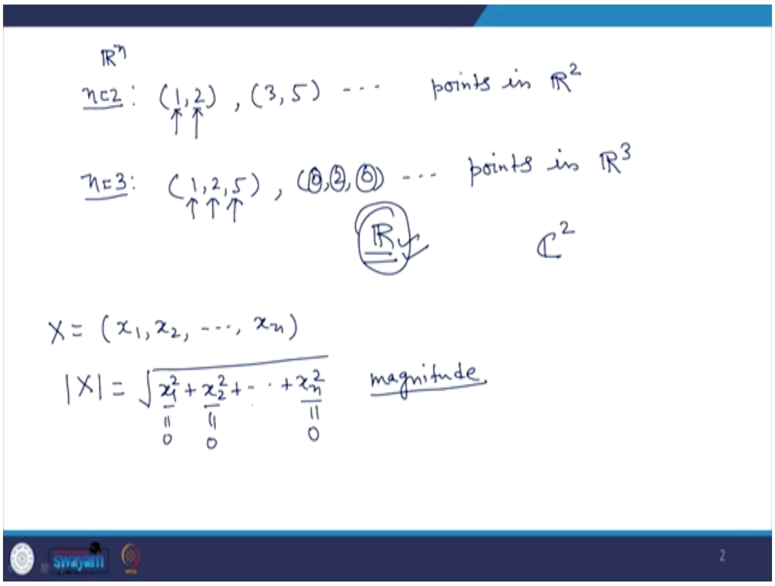
In next 20 lectures which will be covered by my colleague Prof. S. K. Gupta, Shiv Kumar Gupta, so he will take lectures on gradient calculus and optimization. And in the last week, he will take the lectures on probability theory. So, we will try to cover all these topics in these forty modulus, so that whoever do not know what mathematics we are having behind all any machine learning algorithm, they will be able to get the idea of mathematics that is involving machine learning. So, in brief I can say this course is an attempt to provide all these mathematical concepts together with some of their applications in machine learning.

(Refer Slide Time: 08:30)



So, as I told you in first 20 lectures, I will cover all those linear algebra and matrix theory. So, I will start my lecture with the definition of a vector. So, what is a vector? So, for each positive integer n, R n denotes the set of all ordered n-tuples x 1, x 2, x 3, x n, where each x i is a real number. So, these are the real numbers. The element of R n are called n-dimensional real coordinate vector or simply vectors. The number x i in a vector are called the components of the vectors. The elements of R is referred to as scalars.

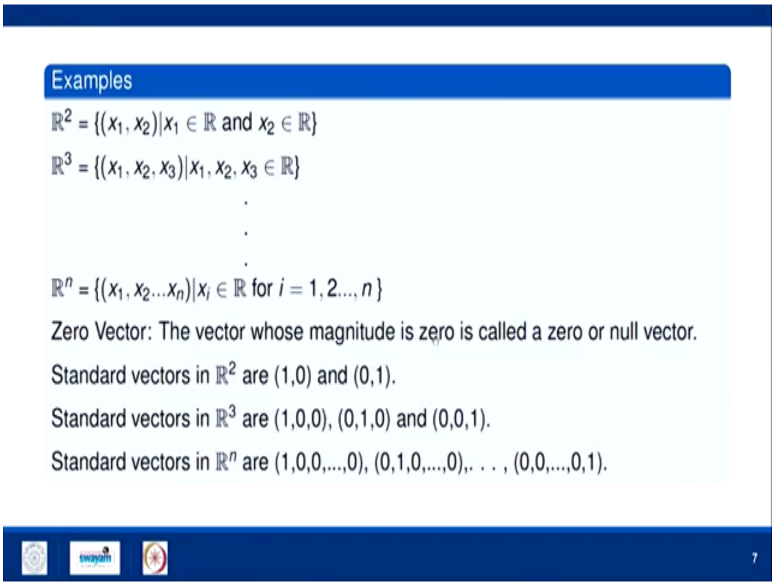
(Refer Slide Time: 09:19)



So, for example, so I am saying here R n. So, if you take n equals to 2, you will be having points like 1, 2, 3, 5, etcetera. So, what is this? This is in points in R 2. So, all the points in R 2 represent vectors, where first component is the x coordinate and second component is the y component. For n equals to 3, we will be having like this 1, 2, 5, 0, 2, 0 and so on, so these are the points in R 3. So, points in three-dimensional space are the vectors of R 3.

So, where first component is x coordinate, second component is y coordinate, and third component is z coordinate and so on. So, all these are the vectors. Individual components are scalars coming from R. So, in this similarly we can have the complex means c 2, c 3 and so on, but this course is more related to real numbers.

(Refer Slide Time: 10:49)

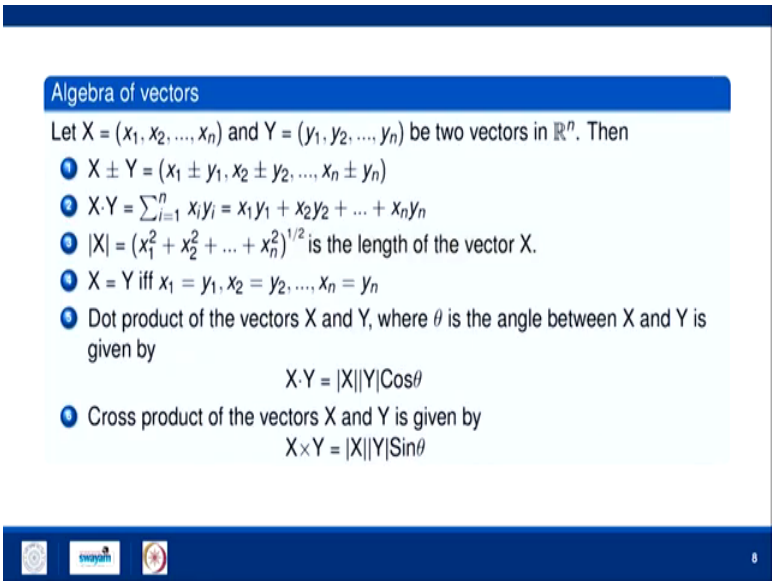


So, zero vector, the vector whose magnitude is 0 is called a zero or null vector. First what we mean by the magnitude of a vector? So, if you are having a vector let us say x 1, x 2, x n, and this is let us say vector X, then magnitude of X is given by this one. And it is nothing just x 1 square plus x 2 square plus x n square means sum of the squares of each component, and then square root of that sum. So, this is the magnitude.

So, if magnitude is 0, then the vector is called zero vector. And when it will happen when all the components are simultaneously 0, then only magnitude will become 0, because it is sum of squares, so only squares should be having 0 value. And when it will be having zero value, when all the components are 0. So, in other way I can say the a zero if all the components of a vector R 0, then the vector is called zero vector.

A standard vectors in R 2 are 1, 0, 0, 1. So, 1 0 representing the x-axis, and 0 1 is representing the y-axis. A standard vectors in R 3 are 1, 0, 0 that is x-axis, y-axis and z-axis. Similarly, standard vector in generalize R n is space are 1, 0, 0, 0, 0, 1, 0, 0, and so on. So, only one component is 1; rest are 0.

(Refer Slide Time: 12:28)

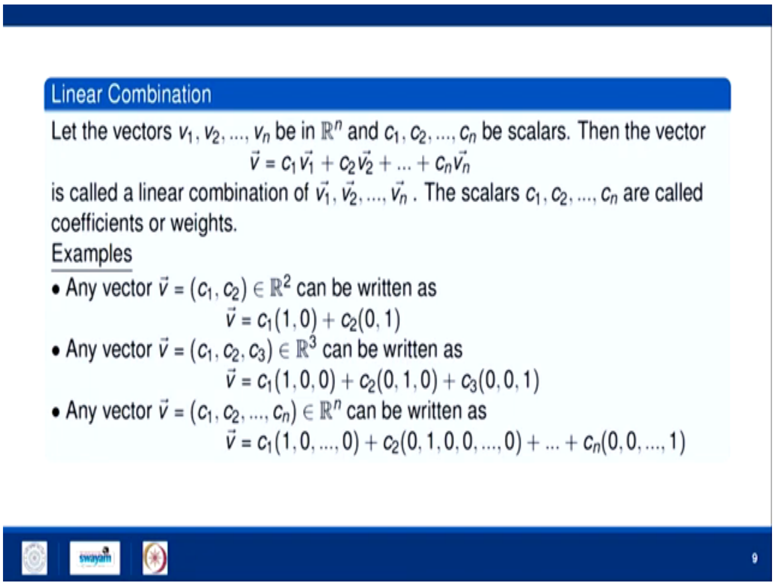


So, algebra of vectors. So, if you are having two vectors in R n let us say capital X which is having component x 1, x 2, x n, and capital Y which is having component y 1, y 2, y n, then x plus y is component y is addition that is x 1 plus y 1, x 2 plus y 2, x n plus y 1. Similarly, x minus y will become first component minus first component of y, second component of x minus second component of y and so on.

The dot product between these two vectors are given by x 1 y 1 plus x 2 y 2 plus x n y n, or I can say i equals to 1 to n, x i y i. Magnitude I have already told you, when we say that two vectors are equal, when all of their components are equal means first component of X equals to first component of Y, second component of X equals to second component Y and so on.

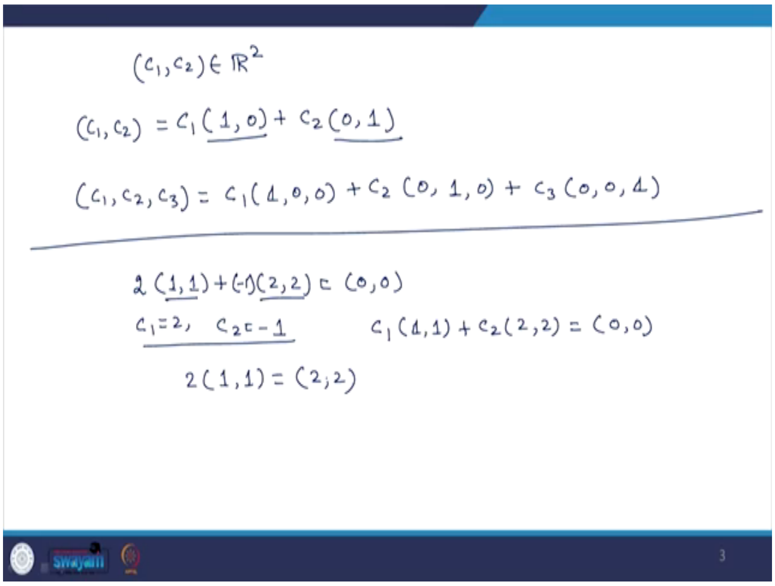
Also the dot product is given by X dot Y equals to magnitude of x into magnitude of y into cos theta, where theta is the angle between X and Y means these two vectors. Similarly, the cross product of the vectors X and Y is given by magnitude of X into magnitude of Y into sin theta again theta is an angle between the vector X and Y.

(Refer Slide Time: 14:05)



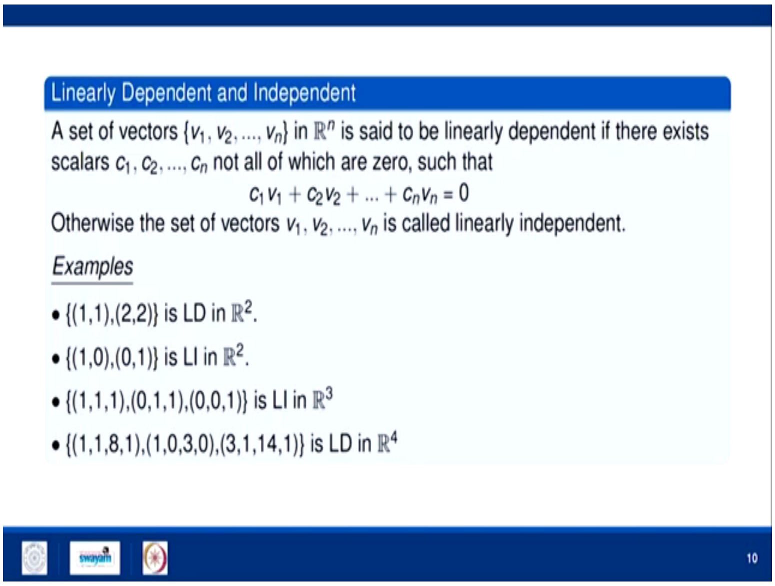
Now, we are defining a very important concept which we will use in many of the subsequent lectures that is called linear combination. So, let the vectors v 1, v 2 and v n be in R n; and c 1, c 2, c n be scalars, then the vector v is the c 1 v 1 plus c 2 v 2 plus c n v n is called a linear combination of the vectors v 1, v 2, up to v n. The scalars c 1, c 2, c n are called coefficients or weights.

(Refer Slide Time: 14:42)



So, for example, if c 1, c 2 is any arbitrary vectors in R 2, then I can write it in linear combination of these two vectors. In fact, c 1 and c 2 are scalars, they are real number. So, all the vectors in R 2 can be written in the linear combination of 1, 0, and 0, 1. Similarly, all the vectors c 1, c 2, c 3 in R 3, I can write in the linear combination of 1, 0, 0, plus c 2 0, 1, 0, plus c 3 0, 0, 1, and so on. So, similarly any vector c 1, c 2, c n of R n can be written as the linear combination of the standard vectors of R n.

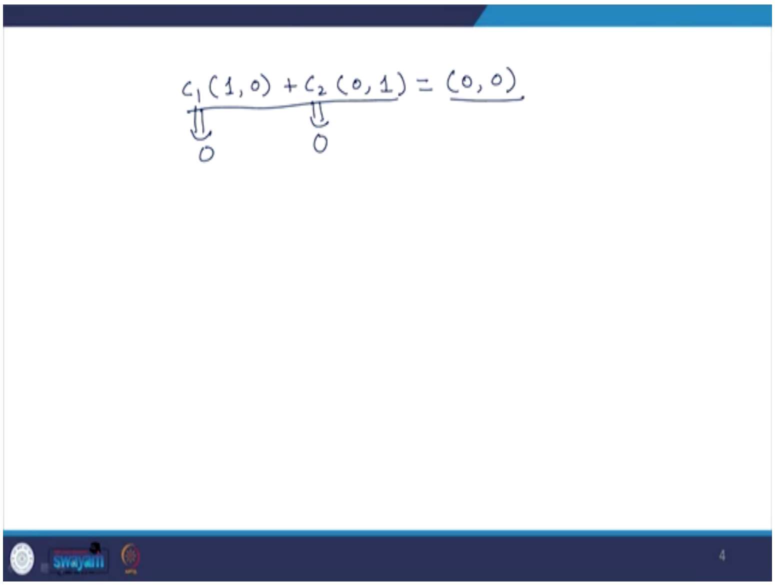
(Refer Slide Time: 15:48)



Now, my next definition is linear dependent and independent. A set of vectors v 1, v 2, v n in R n is said to be linearly dependent. So, all these are vectors in R n. So, this set is said to be linearly dependent if there exist scalars c 1, c 2, c n not all of which are zero, such that their linear combination is 0; otherwise the set of vectors is called linearly independent.

So, for example, 1, 1, 2, 2 is linearly dependent. How you will do it? So, I am having I have to check 1, 1, and 2, 2. So, 1, 1, and then I am having 2, 2. So, if I multiply this by 2, and this by minus 1, what I will get 0, 0. So, I am having c 1 equals to 2, and c 2 equals to minus 1 such that the linear combination of c 1 1, 1 plus c 2 2, 2 is 0 ok. Hence these two vectors are linearly dependent. And how you can say that 2 times 1, 1 equals to 2, 2. So, one vector is twice of the other. So, it means they are linearly dependent.

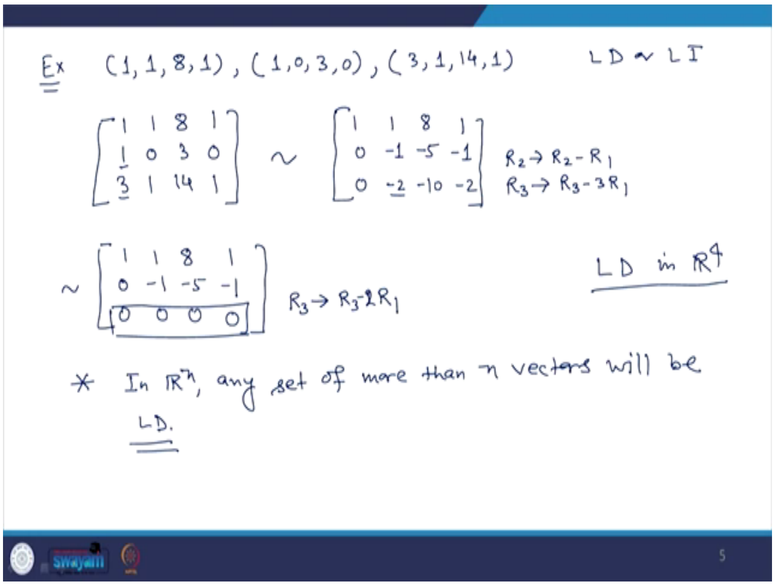
(Refer Slide Time: 17:34)



1, 0, and 0, 1 is linearly independent in R 2. What I want to say that you cannot find such c 1 and c 2 those are non zero or one of them is 0 such that it is equals to 0, 0. So, for having this linear combination equals to 0 vector only when c 1 is 0 and c 2 is 0; otherwise it cannot be. So, hence they are linearly independent.

Similarly, 1, 1, 1, 0, 1, 1, and 0, 0, 1 is linearly independent in R 3. So, how to check whether the given set of vectors is LD or LI that you can find out using the concept of row equivalent form right. Those vectors is the row of a matrix and then apply row ah reduce the matrix in row equivalent form. If you are having all the rows non-zero, then the vectors are LI; otherwise vectors are LD.

(Refer Slide Time: 18:39)



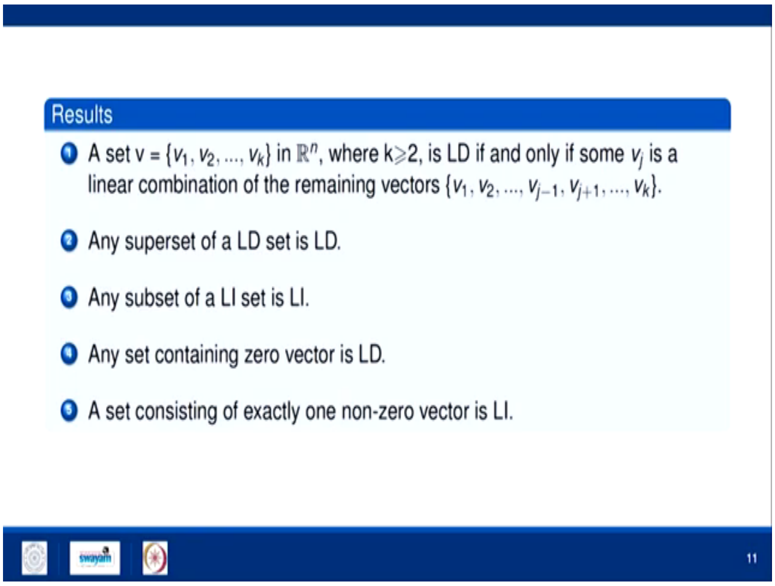
So, for example, take this one 1, 1, 8, 1. So, what I am having 1, 1, 8, 1 1, 0, 3, 0, 3, 1, 14, 1 ok. So, I have to check LD or LI. So, what I will do I will write these vectors as the row of a matrix 1 1 8 1, 1 0 3 0, 3 1 14 1.

So, for reducing it into row equivalent form, I will make this element 0 and this element 0 first. So, what operation I will apply R 2 replaced by R 2 minus R 1. So, I am applying these elementary row operations; and R 3 replaced by R 3 minus 3 R 1. So, no changing first row 1 1 8 1, second row 1 minus 1 0 0 minus 1 minus 1; 3 minus 8 minus 5; 0 minus 1 minus 1. Then third row, 3 minus 3, 0; 1 minus 3, minus 2; 14 minus 24, minus 10; 1 minus 3, minus 2.

Now, I will make this element 0. So, what operation I will apply R 3 goes to R 3 minus 2 times R 1. So, 1, 1, 8, 1 no change in first row; 0 minus 1 minus 5 minus 1; and third 0 minus twice 0 is 0; minus 2 plus 2, 0; minus 10 plus 10, 0; minus 2 plus 2, 0. So, here you can see this row become 0. So, these three vectors are linearly dependent in R 4.

Moreover, one important point to be noticed that in R n, any set of more than n vectors will be linearly dependent, because in R n you can have at most n linearly independent vector. And why, we will see it when I will introduce you the concept of basis.

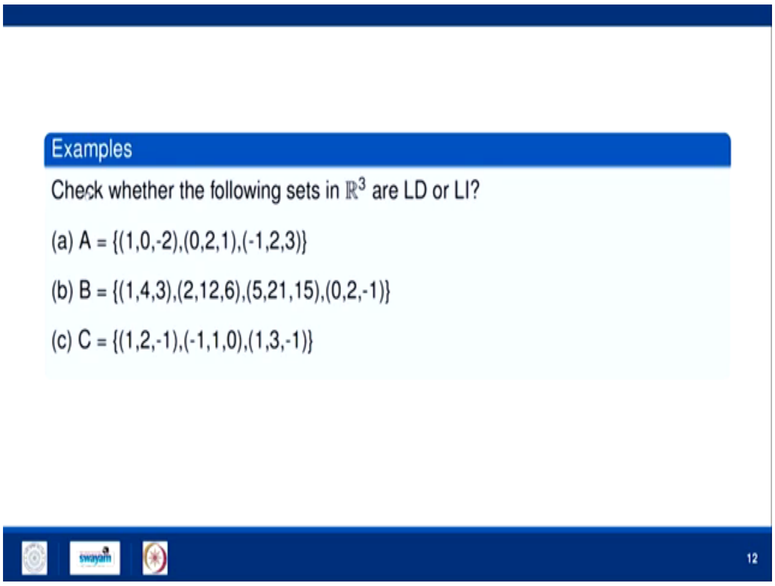
(Refer Slide Time: 21:48)



Now, a set v 1, v 2, v k in R n, where k is greater than equals to 2 is LD if and only if some v j is a linear combination of the remaining vectors v 1, v 2 up to v k means if you can write a vector as a linear combination of remaining of the vectors then we will say that set is linearly dependent. If you cannot do this, then the set will stay or set is to be set linearly independent.

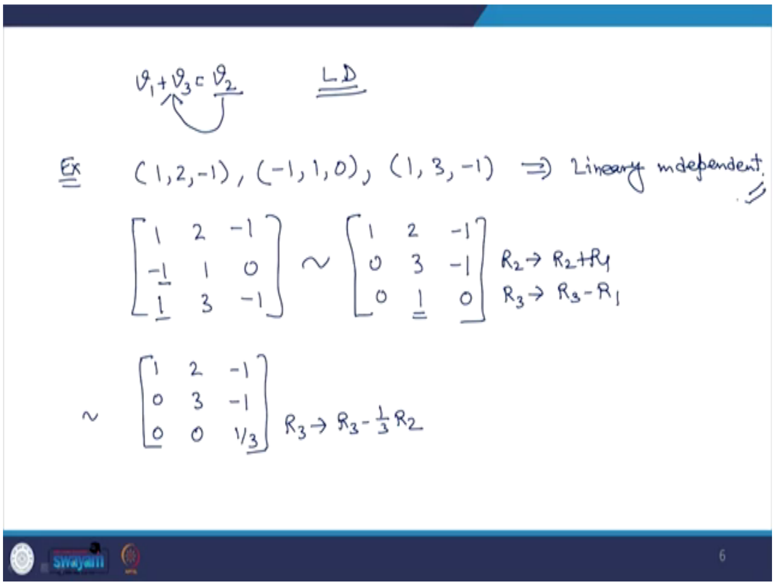
Now, any superset of a linearly dependent set is linearly independent, because already if the set is linearly dependent, you can write a vector as a linear combination of other vectors. So, even if you take the superset this thing will remain there, and the set will remain LD. Any subset of a linearly independent set is linearly independent. Any set containing 0 vector is linearly dependent, because for that 0 vector you can take a non-zero coordinate, non-zero coefficient or non-zero weight in the linear combination or in the definition of LD, LI. So, any set which is containing the 0 vector is linearly dependent. A set consisting of exactly one non-zero vector is linearly independent.

(Refer Slide Time: 23:17)



Now, we are having some of the examples. Check whether the following sets in R 3 are LD or LI. So, first is 1, 0, minus 2, 0, 2, 1, minus 1, 2, 3. So, what I do, if I take the sum of first and third vector, so 1 minus 1 will give you 0, 0 plus 2 is 2, and minus 2 plus 3 is 1.

(Refer Slide Time: 23:48)



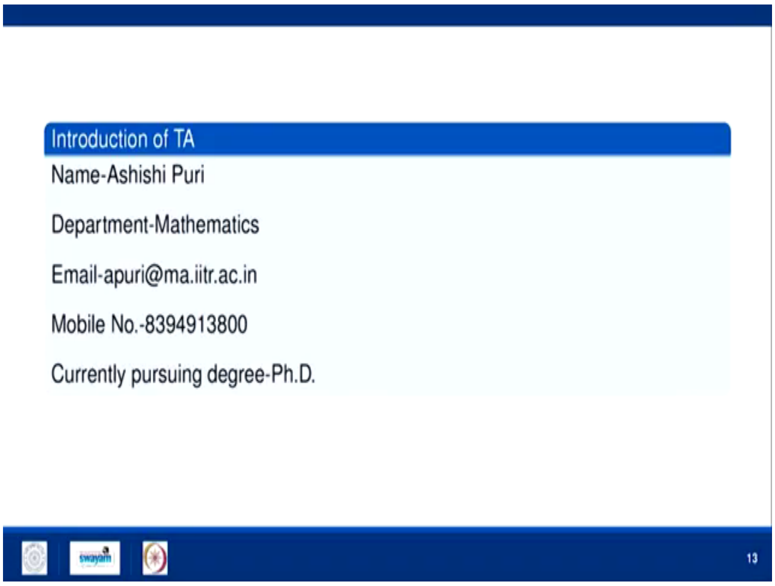
So, what I am having I am having that v 1 plus v 3 equals to v 2 here. So, I can write v 2 as a linear combination of first and third vector. So, this set is a linearly dependent set next one is 1, 4, 3, 2, 12, 6, 5, 21, 15, 0, 2, minus 1. So, these are the vectors in R 3. But here how many vectors we are having we are having four vectors. So, as I told you any set of more than n vectors in n-dimensional vector space is linearly dependent. So, they will be linearly dependent because in R 3 we are having four vectors.

Now, 1, 2, minus 1, minus 1, 1, 0, 1, 3, minus 1, so again what I am having 1, 2 minus 1, my next vector is minus 1, 1, 0, and my next vector is 1, 3, minus 1. So, as I told you use the same matrix method minus 1, 1, 0, 1, 3, minus 1. Change it in row equivalent form, so make the 0 and make the 0.

So, R 2 replaced by R 2 plus R 1, and R 3 replaced by R 3 minus R 1. So, 1, 2, minus 1; minus 1 1, 0; 1 plus 2, 3; 0 minus 1, 1 minus 1 0; 3 minus 2, 1; and minus 1 plus 1, 0; Again make this vector 0. So, what I will do R 3 replaced by R 3 minus 1 by 3 R 2. So, 1 2 minus 1, 0 3 minus 1, 0, 1 minus 1 0, and this will become plus 1 by 3.

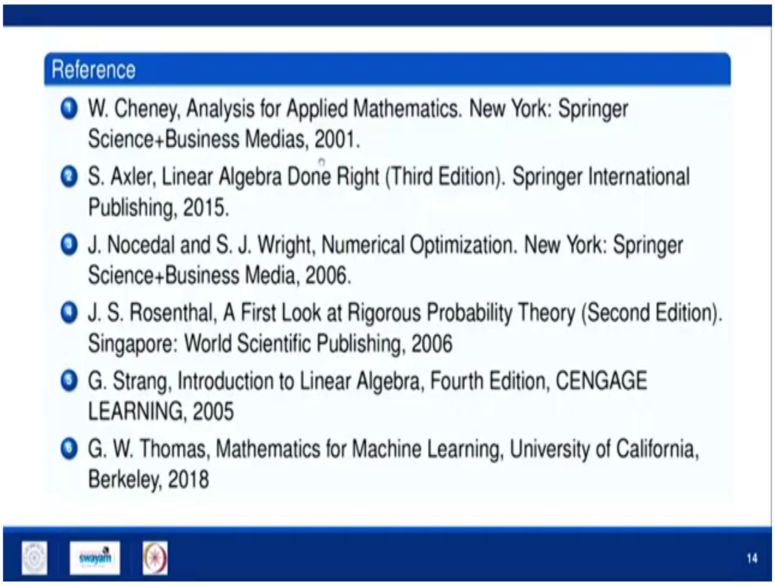
So, you can see all the threes are non-zero and the matrix reduce into row equivalent form. So, these three vectors are linearly independent. So, what we have learn we have learn the concept of linear combination and then we have using that the concept of linear combination. We have given the definition of linearly dependent and linearly independent set of vectors. In the next lecture, we will talk about the concept of vector space, then subspace, and we will go for to the linear transformation.

(Refer Slide Time: 27:06)



For the first 20 lectures Ms. Ashishi Puri will be the teaching assistant in this course, see from Department of Mathematics IIT Roorkee. Her email id is apuri at ma dot iitr dot ac dot in. And mobile number is this one. Currently, she is pursuing Ph.D, here at Mathematics Department IIT Roorkee. This is for first 20 lectures those I will teach you. The next 20 lectures or 20 modules we will be taught by Prof. Shiv Kumar Gupta; and for him Ms. Vrinda will be teaching assistant. So, he will give you detail about the teaching assistant.

(Refer Slide Time: 27:41)



These are some of the references for this course. However, if you go talk about first 20 lectures, then this Gilbert Strang book is nice book, even Thomas notes on Mathematics for Machine Learning that is also very nice. But very short we will discuss many more thing apart from these notes. So, you can follow this book. For the rest of the 20 lectures, you go for this book Numerical Optimization and A First Look at Rigorous Probability Theory. So, this is all about first lecture. Hope you have enjoyed this lecture.

Thank you very much.