Indian Institute of Technology Madras

NPTEL

National Programme on Technology Enhanced Learning

Pattern recognition

Module 01

Lecture 02

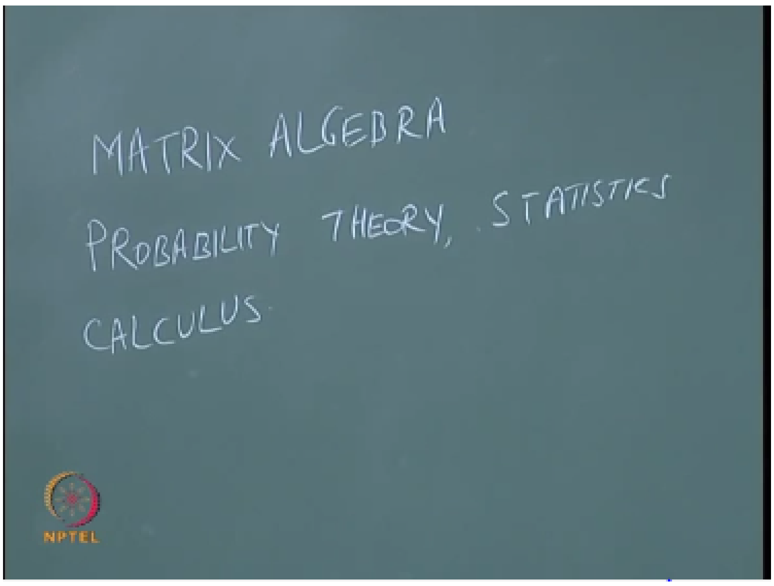
Principle of Pattern recognition II
(Mathematics)

Prof. C. A. Murthy

Machine Intelligence Unit,
Indian Statistical Institute, Kolkata

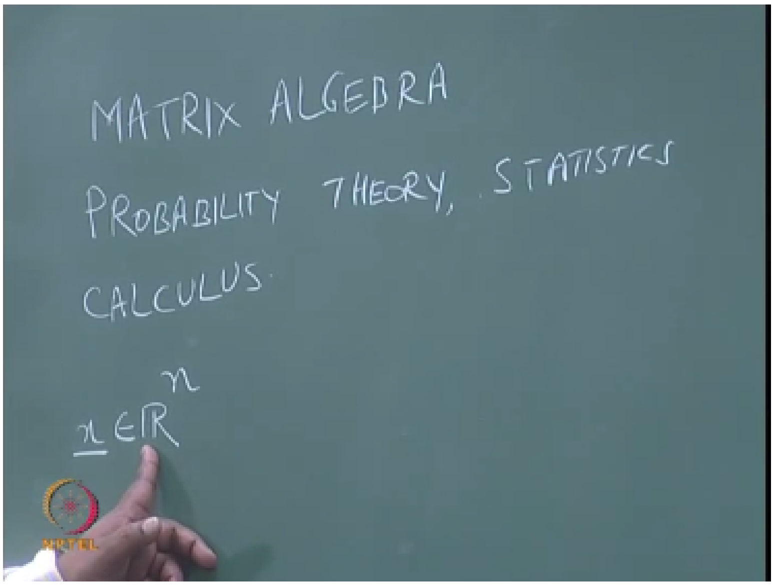
So I have actually given you some sort of an introduction to pattern recognition where I basically mention the uses of pattern recognition, now I shall be going into the mathematics. Before I actually enter into the subject of pattern recognition we need to know a little bit of mathematics, preliminary mathematics, so that we can actually use those things, use the mathematics while developing theorems are other results for pattern recognition. The basic mathematics involves some amount of matrix algebra and basics of probability theory and statistics and yes, I also assume some amount of calculus knowledge, some amount of calculus knowledge.

(Refer Slide Time: 01:36)



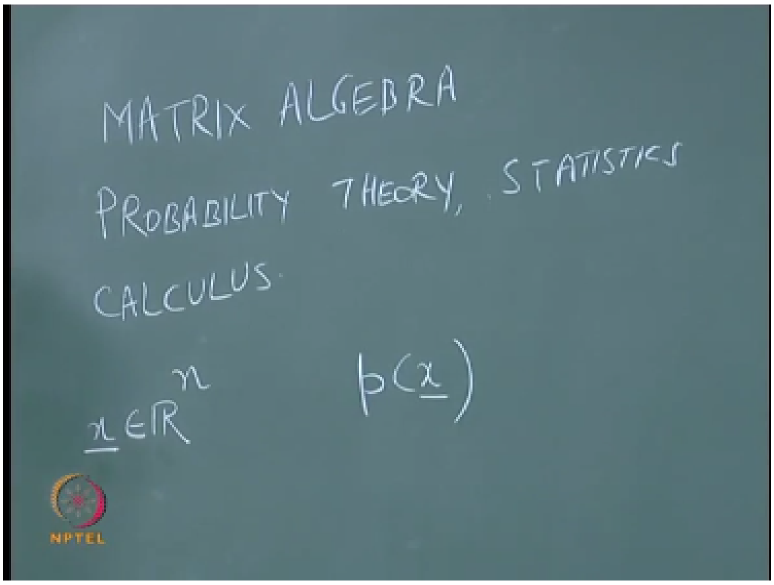
I assume that all of you have this knowledge, I assume that all of you know the meaning of what a probability density function is and I assume that you know what a Gaussian distribution is. Let us just see, a probability density function is, it may be defined over N dimensional Euclidean space which I represented by RN.

(Refer Slide Time: 02:10)



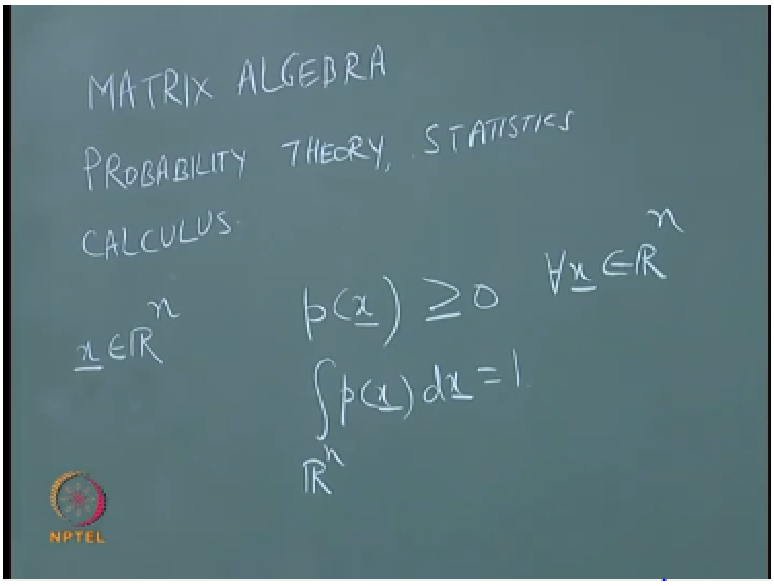
This is a point in n dimensional Euclidean space. This is actually a vector and a column vector, it is a column vector. It has smaller number of components and this belongs to N dimensional Euclidean space and a probability density function P defined over N dimensional Euclidean space.

(Refer Slide Time: 02:44)



P defined over N dimensional Euclidean space, it has the following properties, it is greater than or equal to zero for all x belonging to RN and integral Px dx over RN = 1.

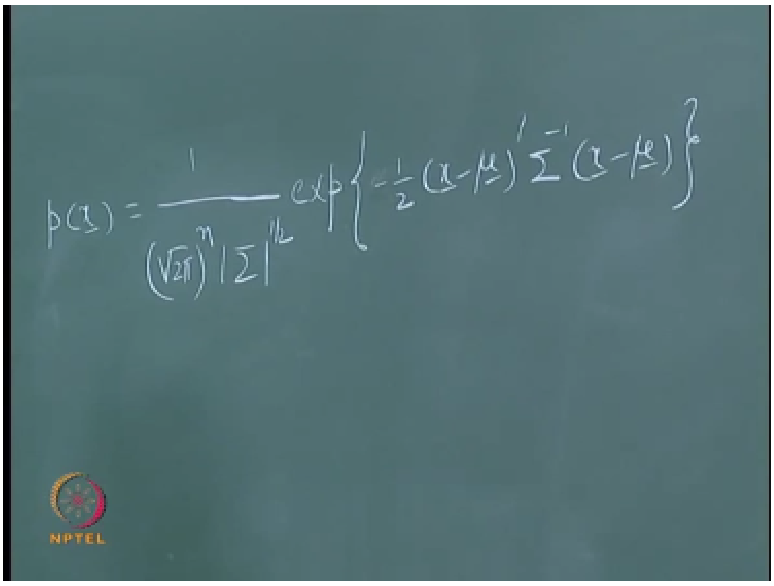
(Refer Slide Time: 03:04)



This is a probability density function over N dimensional Euclidean space. This DX it means actually D the first component if I call it X1 then DX1 DX2 over DXN. So that is just represented as DX and there is this underline you will see at all these places, this is representing column vector. It is representing a vector. And integral PX DX over the whole of RN = 1.

So any function satisfying these two properties is known as probability density function over N dimensional Euclidean space. We can actually in the course of this lectures, we can assume that we are not going into complex spaces, basically we are going to be in the real spaces. So our probability density function is like this and now there are many such functions like this. One such function is known as density function for Gaussian distribution or Gaussian density function. Density function for Gaussian distribution that is also known as normal density function. It is, the definition is this.

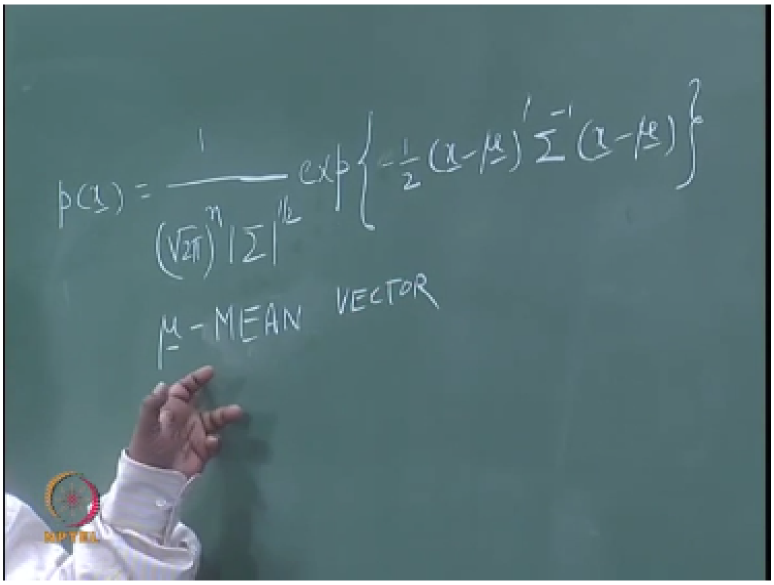
(Refer Slide Time: 05:13)



$(1/\sqrt{2}\Pi^{n})$ determinant of $\sigma^{1/2}$ exponential to the power – ½ x-μ', prime means transpose σ^{-1} x-μ. Here there is a determinant of $\sigma^{1/2.}$ I think since you have got background in matrix algebra, determinant of σ it need not always be positive. It can be zero or it can also be negative. If determinant of σ is negative then determinant of σ to the power half you cannot write because then it becomes a complex number. And our definition of probability density function, not only our definition it says it must be strictly greater than or equal to zero.

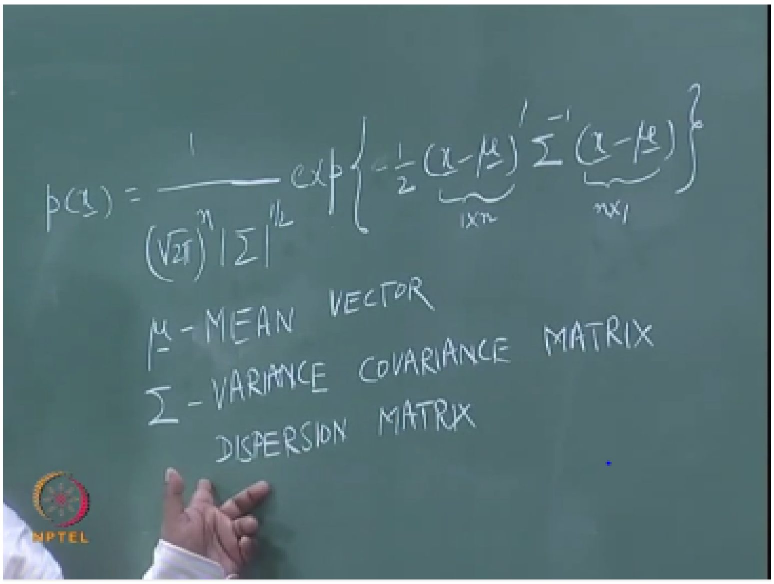
So this means that necessarily this should be greater than zero. The determinant of σ should necessarily be greater than zero. Even if it is zero this whole quantity is not defined. Even if it is zero this whole quantity is not defined. So this should be strictly greater than zero and this exponential means E, E to the power of something. Here let us just see this x is an N dimensional vector, Mu is the mean vector. Mu is what is known as mean vector.

(Refer Slide Time: 07:20)



It is the mean, it represents the mean of the distribution. Mean average of the distribution and this is also small N dimensional vector. X is also small N dimensional vector. Both of them are column vectors, so when I write the transports they become row vectors that means this is going to be 1/N. So here let us look at this, this is x- μ this will be n/1. Now what is now the σ? σ is known as, there are several names for it, it is known as variants covariance matrix. Or another name for it is dispersion matrix, variants covariance matrix or dispersion matrix.

(Refer Slide Time: 08:41)

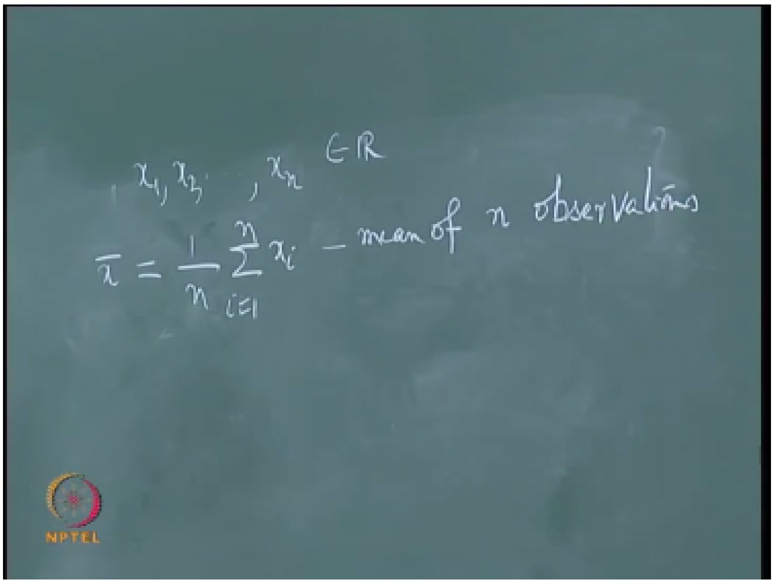


This will be n/N matrix, σ is an n/n matrix so σ inverse will also be n/n. So then this whole thing will be 1/n, n/n, this whole thing will be 1/1 that is the scalar. So e power, so any quantity is greater than or equal to zero, right, e power any quantity is greater than or = 0 and this is scalar so this will be greater than or equal to zero. Anyway I have already mentioned that the determinant of σ has to be strictly greater than zero, determinant of σ has to be strictly greater than zero. Now how does one ensure it, that let me tell you how one ensures that.

First do you all understand the meaning of dispersion matrix variants covariance matrix, do you understand, yes or no? I think I will explain. So I will explain to you what a dispersion matrix is or what a variance covariance matrix is. For this one you need to know first the meaning of variance and you need to know the meaning of covariance, you need to know the meaning of variance as well as you need to know the meaning of covariance.

If you have n observations X1 X2 XN. These are points in R, that means these are n values on real line, then the mean of this n values x bar = I over n, Σi = 1 to n xi, this is the mean of this n observations.

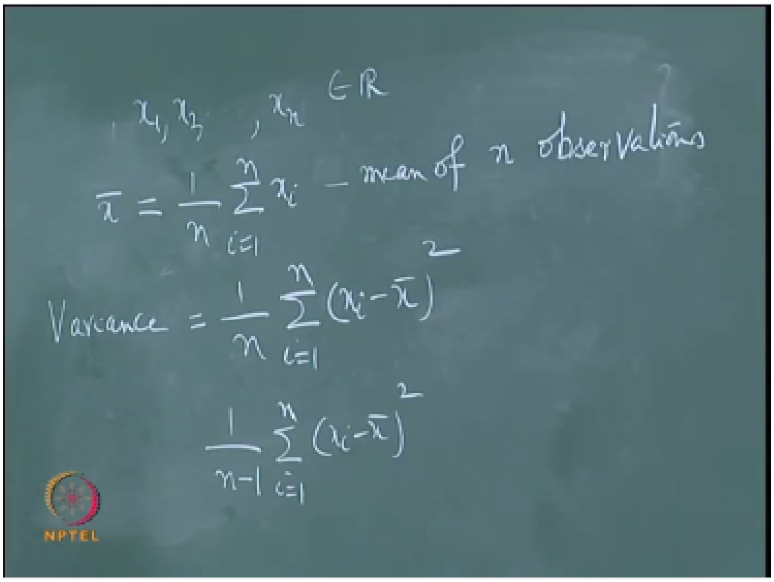
(Refer Slide Time: 11:10)



Then the variance of this n observations, the variance is 1over N Σi= 12N XI-X bar whole squared. The variance is - 1/N Σi=12N, XI-X bar whole squared, that is from every point you subtract the mean and you take the square, like that you do it for all the n points and take the average of this squares, that is 1/n, Σi = 12N, XI-X bar whole squared. In some books you would find some other expression like 1/n-1, Σi=12n xi-x bar whole square, in some books you will find this also. This is actually unbiased estimate for variance of the population.

This is known as unbiased estimate for variance of the population, this is a slightly complicated concept in statistics, let us not going to those specifics there what is n or N -1, let us not go into the specifics, we can just follow one of these things and I would like to follow this.

(Refer Slide Time: 12:56)

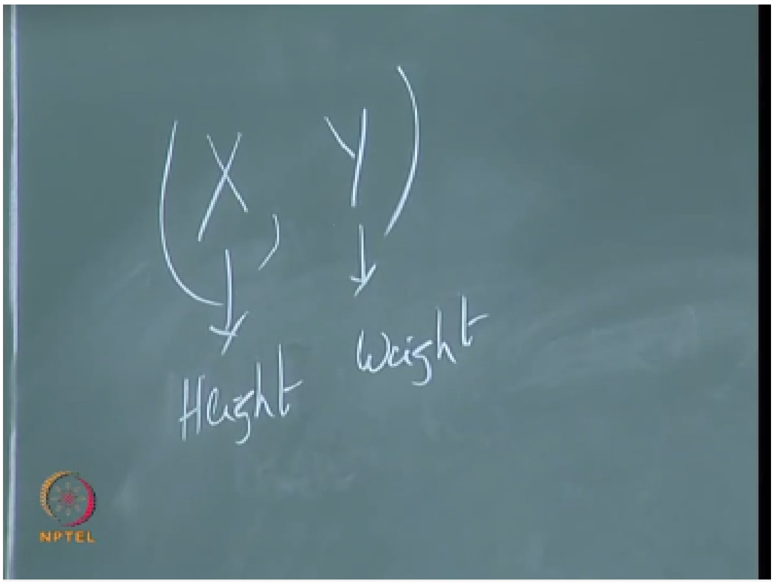


But if you are interested you can also follow this n-1 also. There will be slight differences in the actual values if you follow n-1 instead of n but ultimately for the decisions it is basically looking at the definition of variants in a slightly different way than, if you write 1/n or 1/n-1 basically it tells that you are looking at the concept of variants in two different ways. There is a slight difference, if you just follow one of these things it is fine and I am following 1/n but you can also follow 1/n-1.

1/n-1 also tells you that it is an unbiased estimate of there is something called a population variants for that the unbiased estimate is e 1/n-1 Σi=1 to n (xi-xbar)^2. This basically for advance students of statistics they will use 1/n-1. The advance students of statistics they will use 1/n-1. So I am going to follow 1/n which is what generally in the preliminary level 1/n Σi=1 to n (xi-xbar)^2 is not in the preliminary level, in the advance level it becomes 1/n-1. This is the basic difference. Now this is about variants. Now there is another concept that is covariance. What is covariance?

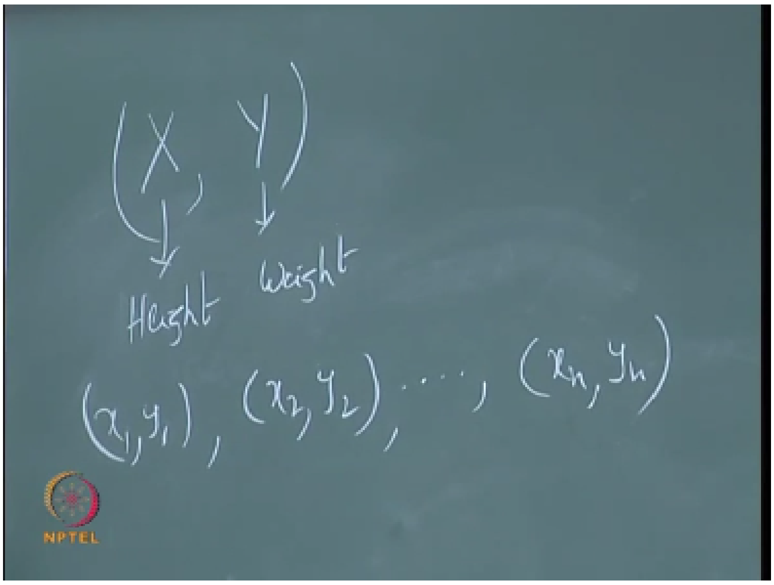
In order to explain the meaning of the word covariant we need to have two variables, let us write the two variables as x and y, it is something like x is say height, y is say weight.

(Refer Slide Time: 45:17)



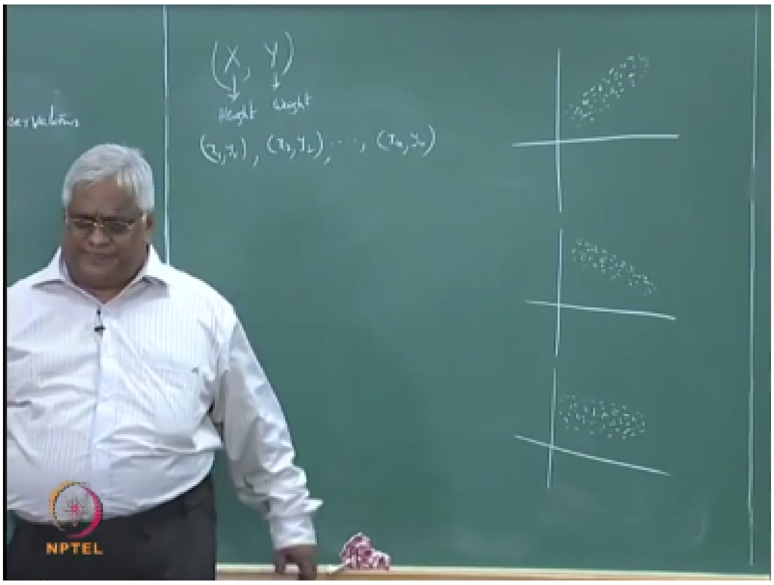
On the same individual you are measuring the individuals height say in cm, individuals weight say in kg. Like that you are measuring these heights and weights say for smaller number of individuals, then you are going to get observations like x1 y1, x2 y2 and you will get xn yn.

(Refer Slide Time: 45:57)



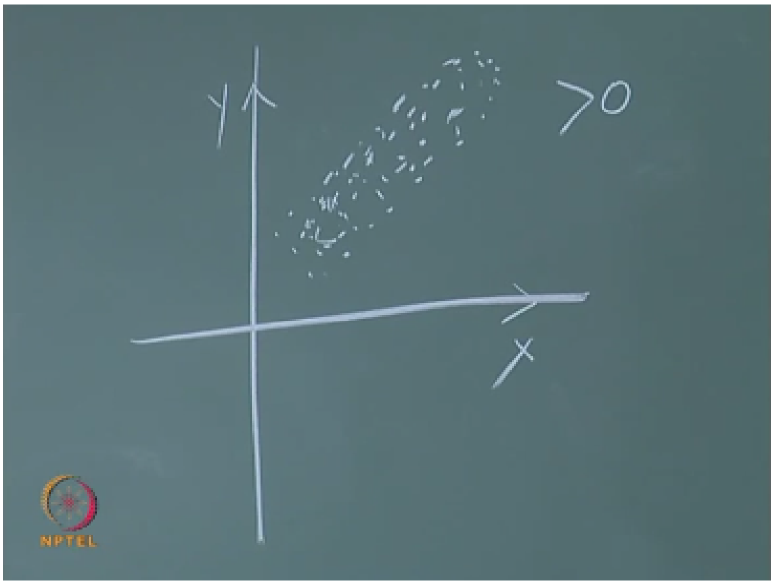
Now what you do is that you just plot these n values. Now I shall give here three such plots. Actually these plots can look in very many ways; I am going to give examples of three such plots. In one plot the points are looking like this. In another plot that is here the x and y variables are such that xn and yn, xi = 12n, yii= 12n. So this xi and yi they are looking like this. In the third plot, xi and yi = 12n, those points are looking like this.

(Refer Slide Time: 47:37)



Now let us see, here this is say your x's, and these are your y's. Here as x values are increasing the y values are also more or less increasing. So we would like to denote this relationship by some quantity that is greater than zero. We would like to denote this relationship by some quantity greater than zero.

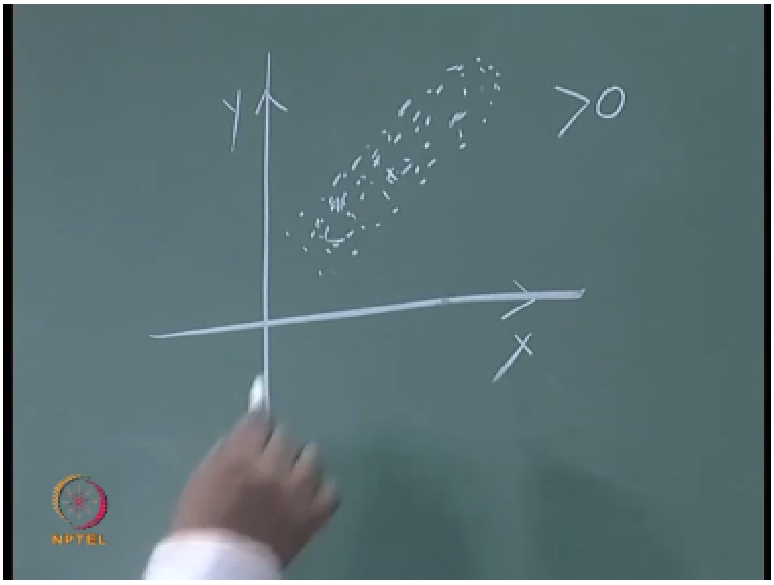
(Refer Slide Time: 18:17)



Now let us look at this, here when x's are increasing y is decreasing. So this is a negative relationship. Here we would like to represent this relationship by the same quantity but that should be less than zero. Now let us look at this, here whatever may be the value of x, y is more or less in the same range. So in this case we would like to get the relationship as something very close to zero. In this case we would like to get as something very close to zero.

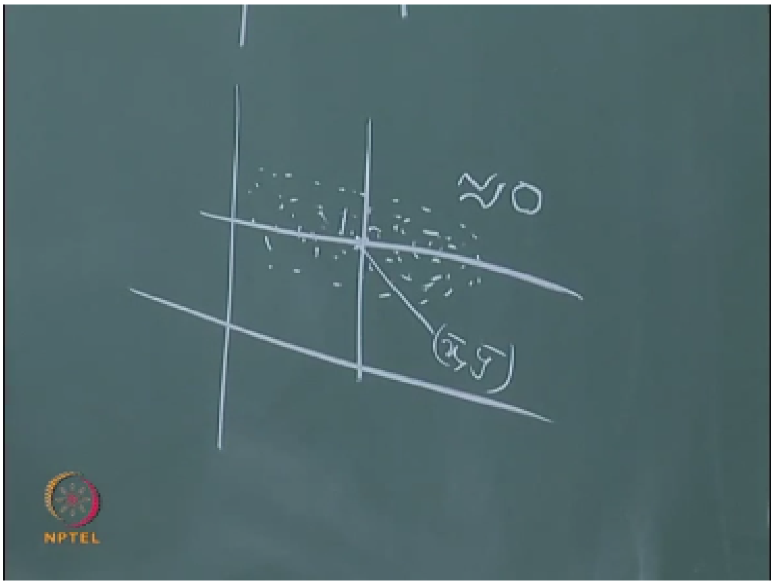
So we would like to define a quantity in such a way that quantity should take positive value here, negative value here. It should be something very close to zero here. What is that quantity? Let us see what that quantity is. For all the exercise here you find the average of these x's, and you find the average of these yi's also. So probably that point will be somewhere here.

(Refer Slide Time: 19:52)



So then what I will do is that I change my axis to this. So this point is x bar y bar. Similarly the x bar and y bar here for this quantity is x bar y bar, for this set of it, and for this set probably x bar and y bar is this, so this is your x bar and y bar.

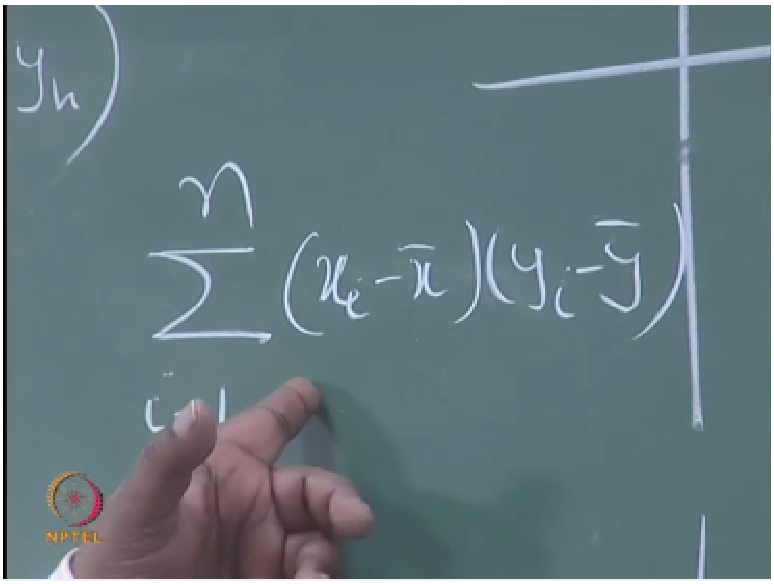
(Refer Slide Time: 20:42)



Is this clear? Now what I will do is that take a point, let us look at the coordinates of this point with respect to new x bar and y bar then at this corresponding value for x is this and the corresponding value for y is this, right. Similarly, for a point here the corresponding value for x is this and y is this. For a point here, the value for x is this and y is this and then so on. Now you multiply the new x and y values, then what is going to happen? Note that this is the first quadrant according to the new axis, this is the second quadrant, this is the third quadrant and this is the fourth quadrant.

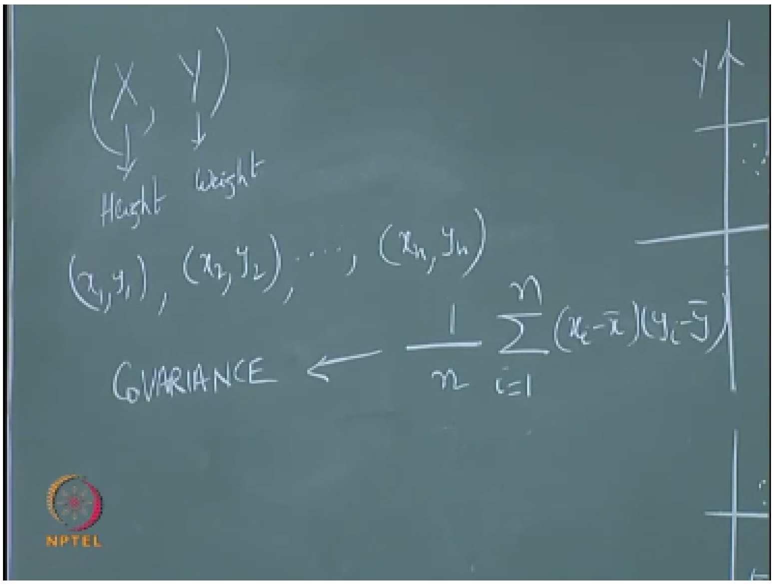
Here x's and y's they are greater than 0, product will be greater than 0. Here x's and y's are less than 0 product will be greater than 0 and here the product will be less than 0. Note that the place where products are less than 0, the number of such points is small whereas the number of points for which the product is greater than 0 that is large and the values are also large, so if you have Σi=12n, xi-x bar x yi=y bar.

(Refer Slide Time: 22:45)



Which is actually this into this at this point, this into this at this point and then soon and then you are just adding them up and again you can have 1 by n or 1/n -1. I am taking 1/n here. This will be greater than 0 in this case, right. Now what will happen here? Here the points are in the second and fourth quadrants, so the product will be less than 0, so here this quantity will be less than 0. What will happen to this? Here the points are more or less equally distributed in all the quadrants, so this Σ is likely to be very close to 0, this is known as covariance, okay.

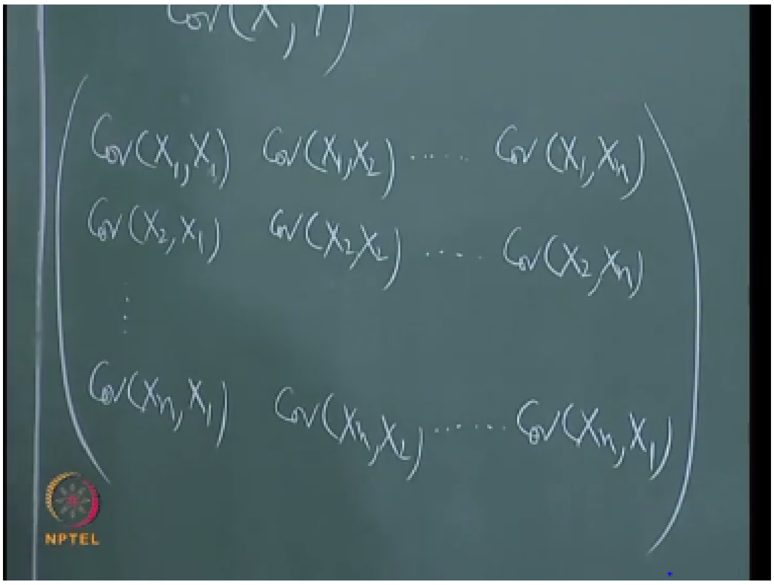
(Refer Slide Time: 24:04)



This is known as covariance between x and y, okay. This is also represented as cov, you only write the variables x, y. Now if you have two variables your covariance system, I said that we have points in N-dimensional Euclidean space that means basically we are assuming that we have smaller number of variables which in pattern recognition language we call them as features, okay. We have smaller number of features or smaller number of variables. So if you take pairs you have how many pairs? You have nc2 pair's right. By the way, what is covariance of x with x, it is basically the variants, is this clear?

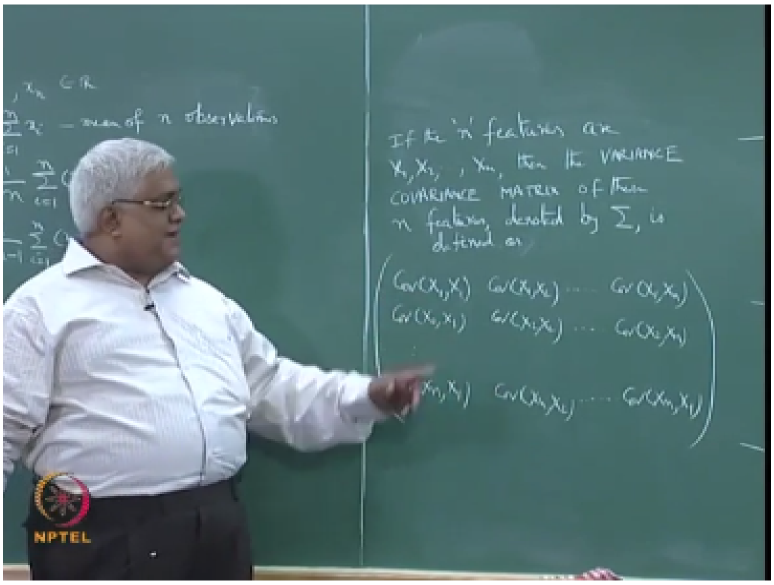
Covariance of x with itself that means (xi=x bar x xi-x bar, xi-x bar)^2 which is what in the variants here, okay. So covariance of x with itself is basically the variants, okay. So if we are in small dimensional space that means basically we have the number of figures are smaller than the variants, covariance matrix is like this. This is the variants, covariance matrix.

(Refer Slide Time: 27:00)



If the n features are x1, x2, xn then the variance, co-variance matrix of this n variables or n features. These n features this is denoted by sigma is defined as this. There are smaller numbers of variables; I am representing number x1, x2, xn since we are in the n-dimensional Euclidean space. There are smaller numbers of variables or smaller number of features and those variables are represented by x1, x2, xn then the variance co-variance matrix is this.

(Refer Slide Time: 29:02)

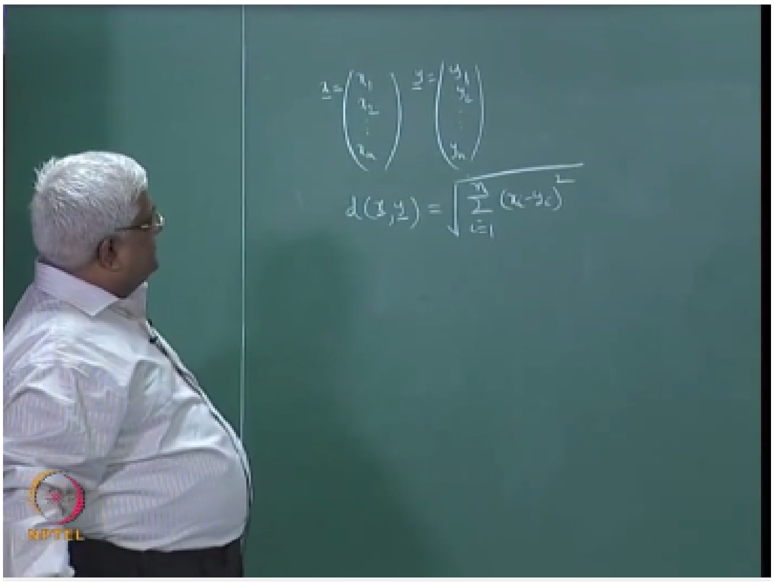


This is an n/n matrix it has n rows and n columns. This is an n/n matrix. The first quantity here is co-variance of x1 with x1, which is nothing but variance of x1. This is co-variances of x1 with x2. If you look at the definition of co-variance whether you write x1 with x2 or x2 with x1 they are the same thing, so this co-variance of x1 with x2, this is x1 with x3, x1 with xn. This is x2 with x1 or x1 with x2, this is x2 with x2 and then so on.

So since co-variance of x1, x2 is same as x2, x1, x1, x3 is same as x3, x1, this matrix is symmetric it is a real matrix, it is a symmetric matrix and it is also what is known as positive definite. It can be shown to be always non-negative definite and in most of the applications it is positive definite. I will explain to you the meaning of positive definite. Probably all of you know the meaning of the word distance. All of you probably also know the meaning of Euclidian distance.

Suppose you have 2 vectors x1, x2, xn this is one vector, y1, y2 yn this is another vector then the distance between these two vectors, if I represent this vector by x, if I represent this vector by y then the distance between these vectors x and y is equal to $(\Sigma i\text{-}12n\ xi\text{-}y)^2$ and then there is a $\sqrt{}$, right. This is the Euclidian distance which all of us know, okay.

(Refer Slide Time: 21:31)



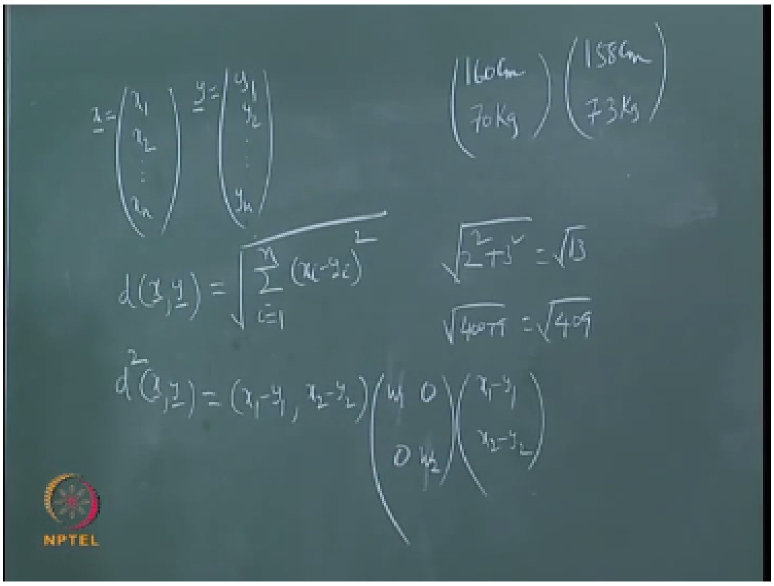
Now let me ask you a question, my question is like this, okay, let us assume that the value of small n is equal to two. We only have vectors in our two dimensional space and let us assume that the first one is representing height and the second one is representing weight. Now say for a person the height let me just say 160 cm, say the weight is 70 kg this is for one person. Say for another person, the height is 158 cm, say the weight is 73 kg, now you want to measure the distance between these two. Now would you like to apply this formula for it? You will have difficulty.

If you want to apply this formula then please note that you are going to have difficulty. What is the difficulty? The difficulty is that if you want to apply this formula one is say i=1, you are going to have the difference between there is a cm value here, there is a centimetre value, 160-158 this is 2, you will get 2^2 and here the next one is 3^2, right. Say I measure height in cm, weight in kg. My friend here, he may want to measure height in mm then what is going to happen? This will be 1600, this will be 1580, the difference will be 20 and this will be 20^2 400, then there is a difference.

I mean originally I got 2^2+3^2, which is $\sqrt{13}$ whereas if I measure this in mm then it is going to be 400+9=409, whereas the two human beings they are the same, so the distance should not change just because I changed the units. The distance should not change just because I changed the units. Always Euclidian distance is not useful. Are you understanding it? Always the Euclidian distance is not useful, then how does one measure the distance, let us see.

Let me write this thing as D^2 xy, so I am just removing √ here this I will just write it as x1-y1, x2-y2, this is a row vector. I am writing 1001 and I am writing this one x1-y1, x2-y2, I am writing 1001 then x1-y1, x2-y2. Do you think the product of this thing is actually Σi= 1 to 2 (xi-y^2) multiplication by identity matrix does not change anything, right. So it is actually going to be Σx-y whole square. Now I said that just because I changed the units the value should not change, right. So now a slight generalization of the distance e instead of 1001 probably we write some weight w1 and w2.

(Refer Slide Time: 26:42)

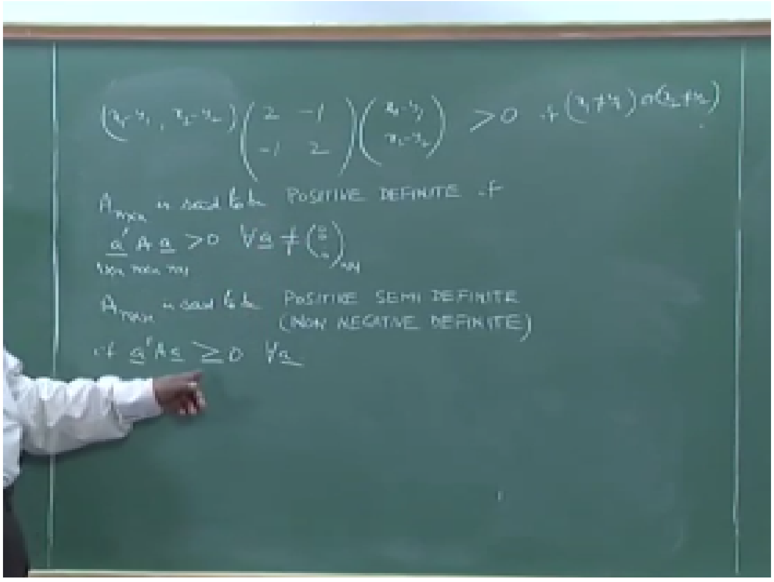


So that here I write wi this depends on the unit. The wi values will change if the units change, so that the whole thing will remain the same. So this is a small change in the definition of distance and you may also have something more instead of this, we might have w11, w12 this is for the present moment let me just take w22, this is a better generalization of the previous one. In the previous one I wrote this, now I want this thing to give us distance, distance by definition it should be greater than or equal to 0.

Now my question to you is for what values of w1 and w2 this thing will be greater than or equal to 0. Now if w1 is strictly greater than or equal to 0 and w2 is strictly greater than or equal to 0 then whatever may be the values of this w1 and w2 if they are strictly greater than or equal to 0 then the whole thing will be greater than or equal to 0, because this whole thing is nothing but (wi and xi –yi)^2 and wi are greater than equal to 0 though the whole sum is greater than or equal to 0, but then if I write a matrix like this, note that I wrote basically a symmetric matrix here w11, this w12 I wrote the same thing here.

And this w22 and this is a generalization of this, okay then for which such matrices this thing will be greater than or equal to 0? I will give you one example, in fact you are going to get many, many such examples.

(Refer Slide Time: 39:40)



Let me just say this, this will be always in fact I will say it is strictly greater than 0. If (x1 is not equal to y1) or (x2 is not equal to y2), you take any x1, y1, x2, y2, if one of these properties is satisfied then this is strictly greater than 0, you can check in your home, any x1 and y1, x2 and y2. One these properties should hold, either this hold of this hold if you both of them hold then there is no problem. Now, I am going to write the definition of positive definite matrix An/n is said to be positive definite if this is an n/n matrix.

If A' AA is strictly greater than 0 for all A0=0 vector, this is the 0 vector. It has n rows and one column. The matrix A is set to be positive definite if A' AA, A is n/1, so A' is 1/n, A is n/n then this n/1, so the whole thing is a scalar, so I can write greater than 0, equal to 0, or less than 0. Now, this is matrix A is said to be positive definite. If this is greater than 0 for all A0=0 vector, and there are many, many such matrices.

And the variance, covariance matrix is what is known as let us just see, alright another definition A n/n is said to be here it is written positive, here it is written as positive semi-definite. In some books, this is also written as non-negative definite. So, matrix A is said to be positive semi-definite or non-negative definite. If A' AA is greater than or equal to 0 for all A. If this is greater than or equal to 0 for all A, and this is strictly greater than 0.

Usually in matrix algebra, you would basically find this definition, but you would not know or generally we may not be knowing why this definition is necessary. I mean a way of looking at this definition is from the point of view of the distances. We want something like this because we wanted this distance to be greater than 0, right? We want this thing to be greater than 0. We want this distance to be equal to 0 if this is equal to this and this is equal to this, then we want the distance to be equal to 0.

Otherwise, we wanted to be strictly greater than 0, right. Otherwise if x1=y1 and x2=y2 we want the distance to be equal to 0, otherwise, we want the distance to be strictly greater than 0. If two quantities are same, then there is no distance. When there is a difference and there is a distance, right. So that is what we want to incorporate in the definition that is why it is written in like this. We want some such definition and such a matrix is known to be positive definite whereas if you include equality and this non-negative definite.

The variance and covariance matrix is, it can be shown to be non-negative definite. The variance, covariance matrix can be shown to be non-negative definite. In fact, most of the times it is positive definite and in the case of normal distribution, we assume that the variance, covariance matrix is positive definite that is why we write in the denominator determinant of $\sigma^{1/2}$. If variance, covariance matrix is positive definite then some properties follow automatically.

What are the properties? Probably, you are aware that the determinant of the matrix is product of its eigenvalues. Now, if it is variance, covariance matrix, then all the eigenvalues because it is non-negative definite they are to be strictly greater than or equal to 0, and if it is positive definite then every eigenvalue is strictly greater than 0.

So that the product is also greater than 0, so that you can write the √ determinant $\sigma^{1/2}$. So, variance, covariance matrix is positive definite implies every eigenvalue of the matrix is strictly greater than 0, okay. So, this is one property that people have used extensively in the literature on pattern recognition. Shall we stop here.

End of Module 01 – Lecture 02

Online Video Editing / Post Production
M. Karthikeyan
M. V. Ramachandran
P. Baskar

Camera
G. Ramesh
K. Athaullah
K. R. Mahendrababu
K. Vidhya
S. Pradeepa
D. Sabapathi
Soju Francis
S. Subash
Selvam
Sridharan

Studio Assistants
Linuselvan
Krishnakumar
A. Saravanan

Additional Post – Production
Kannan Krishnamurty & Team

Animations
Dvijavanthi

NPTEL Web & Faculty Assistance Team
Allen Jacob Dinesh
Ashok Kumar
Banu. P
Deepa Venkatraman
Dinesh Babu. K.M
Karthick. B
Karthikeyan. A
Lavanya. K
Manikandan. A
Manikandasivam. G
Nandakumar. L
Prasanna Kumar. G
Pradeep Valan. G
Rekha. C
Salomi. J
Santosh Kumar Singh. P
Saravanakumar. P
Saravanakumar. R
Satishkumar. G
Senthilmurugan. K
Shobana. S
Sivakumar. S
Soundhar Raja Pandian. R
Suman Dominic. J
Udayakumar. C
Vijaya. K.R
Vijayalakshmi
Vinolin Antony Joans

Administrative Assistant
K.S. Janakiraman

Principal Project Officer
Usha Nagarajan

Video Producers
K.R. Ravindranath
Kannan Krishnamurty

IIT Madras Production

Funded By
Department of Higher Education
Ministry of Human Resource Development
Government of India

www.nptel.ac.in