NPTEL

NPTEL ONLINE CERTIFICATION COURSE

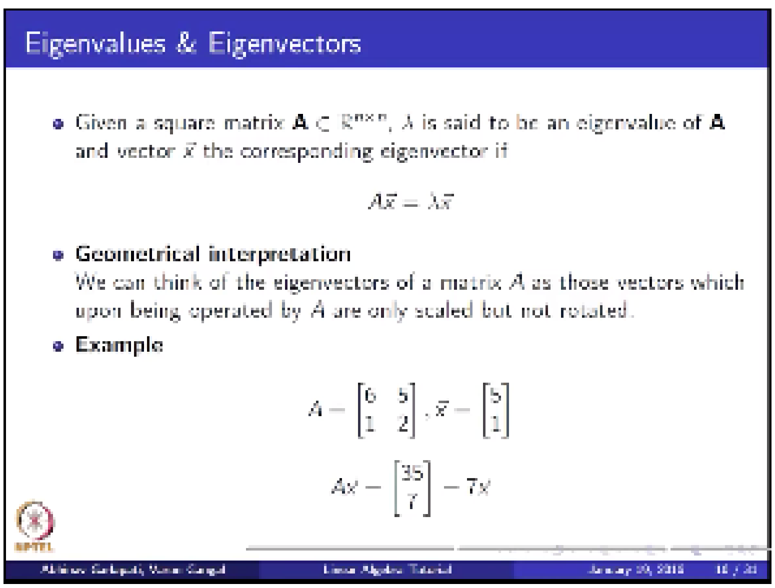
Introduction to Machine Learning

Lecture 6

Prof. Balaraman Ravibdran
Computer Science and Engineering
Indian institute of technology

Statistical Decision Theory – Classification

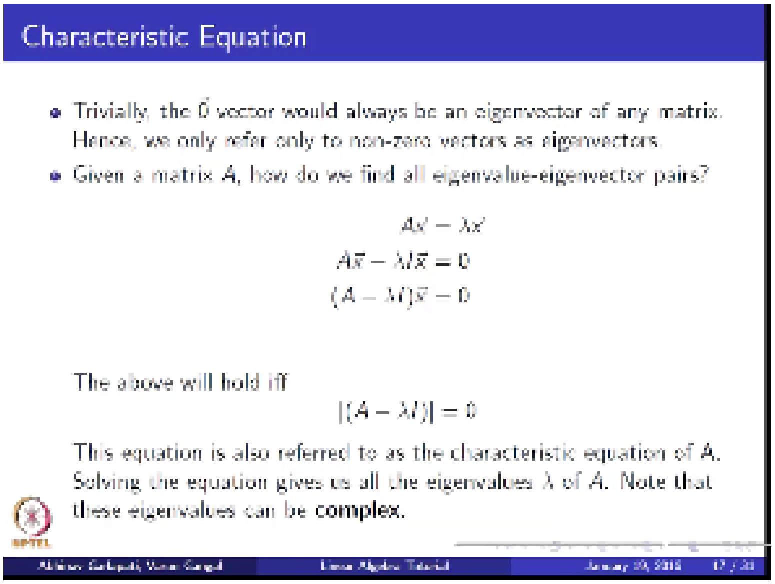
(Refer Slide Time: 00:15)



I am Victor and eigenvalue of are spectively so note here that I in vectors and eigenvalues are tied together which means that every Eigen vector has an Associated Eigen value. We often characterize square matrices inters of their eigenvectors one way of looking at high in vectors is as follows can be thought of as a vector and in our heart for N and the square matrix acts like an operator which transforms Into another n-dimensional vector ax now the Eigen vectors of a are those vectors which on being transformed by a or operated upon by a I am only scared my laptop but not rotated in other words their direction does not change.

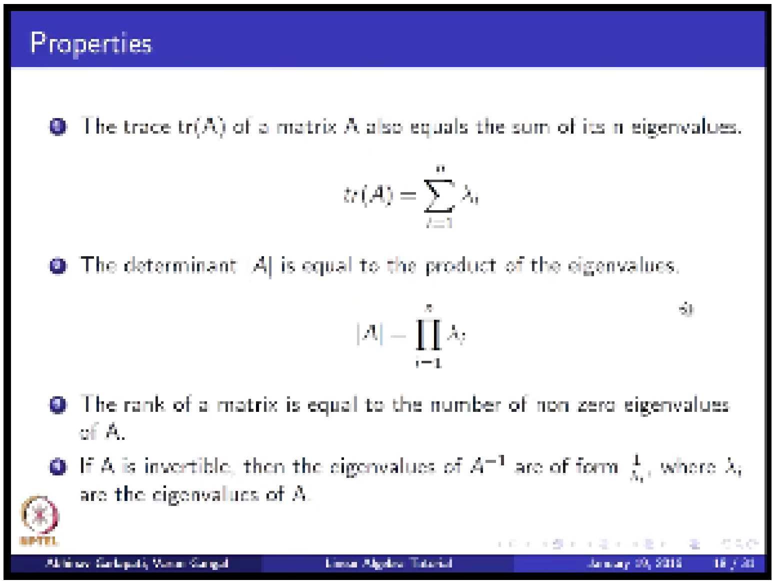
We can have a look at this example here the 2 cross 2 matrixes a on multiplying the vector X /1 gives back the vector multiplied by the real value 7. So when is an eigenvector of a and 7 is aneigenvalue of A.

(Refer Slide Time: 01:40)



We can see that you know would always be an eye effectors of any matrix if we simply go by the ax equals λ X definition hence we only revert nonzero vectors as I cannot read this. So the question is given a matrix a how does one find all the Eigen value eigenvector pairs by simplifying ax equals λ X we get a _ λ I 2 =0 now since we are only looking at nonzero vectors in no fix cannot be 0and expand be a zero vector which means that picked up a - laughter should pay zero so the equation that of a _ λ equals zero is called a characteristic equation of a so I think this equation gives us all the in values of a one thing you use notice that even though all the values of a are real is a real matrix the Eigen values can be complex.

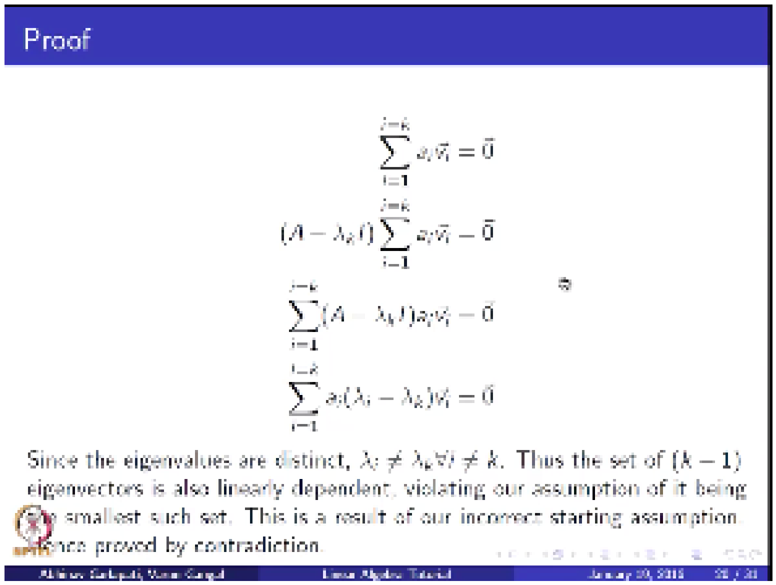
(Refer Slide Time: 02:54)



There are interesting relations between some properties of matrix and its Eigen values for instance the trace of a matrix is equal to the sum of its eigenvalues while the determinant is equal to the product the rank of a matrix is equal to the number of nonzero.eigenvaluesnote that if an eigenvalue has multiplicity greater than 1 for instance if two distinct Eigen vectors x1 and x2both have Eigen value λ.

We would count λ twice also we can describe the eigenvalues of a inverse in terms of the Eigen values of a provided of course a is invertible the eigenvalues of a inverse will be of the form 1 by λ I where λ I is an Eigen value of A.

(Refer Slide Time: 03:55)



.

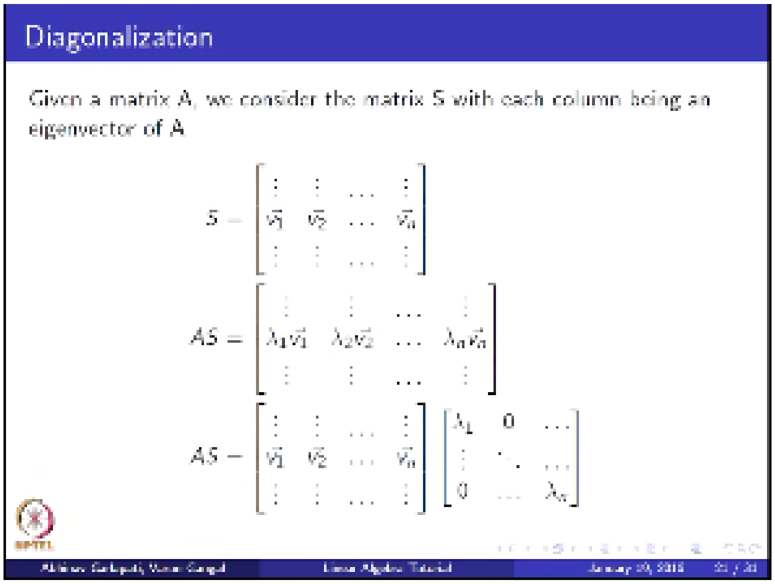
Now let us have a look at an interesting theorem about eigenvalues and eigenvectors the theorem goes as follows if a matrix has all its eigenvaluesdistinct in its Eigen vectors are linearly independent we shall prove this by what is called a proof bycontradictionif this theorem does not fool that means there is a set of K Eigen vectors such that it is linearly dependent let the is vector in the set B be I and the corresponding eigenvalue be λ .

Note that we are considering the smallest such set since the set is linearly dependent this means there exists real called shrimps arises that summation AI VI equal to zero now let us multiply both sides of the equation by a _ λ K times I since VK is an eigenvector of a _λ K IV K will be equal to zero we can understand this from the characteristic equation hence the term corresponding to VK disappears from the equation since it goes to zero.

Now for the remaining Eigen values since we know they are distinct the Tom λ I _λ K cannot be equal to zero note that a _ λ I into V I simplifies to λ I _ λ K times V I since a VI equals λ IVI forever now we can think of a I times I _ λ K as a new constant P I this means now that we have a summation running from I equals 1 to I equals K _ 1 such that B IV I equal to 0 however we had assumed that this is the that the set of size K was the smallest set of linearly dependent I converters however now we have an even smaller set this contradicts our starting assumption

Hence such a set of K linearly dependent I can because cannot exist for any K greater than equal to 2hence all our Eigen vectors are linearly independent hence our theorem stands to diagonalization gives us a way of representing a matrix.

(Refer Slide Time: 06:54)

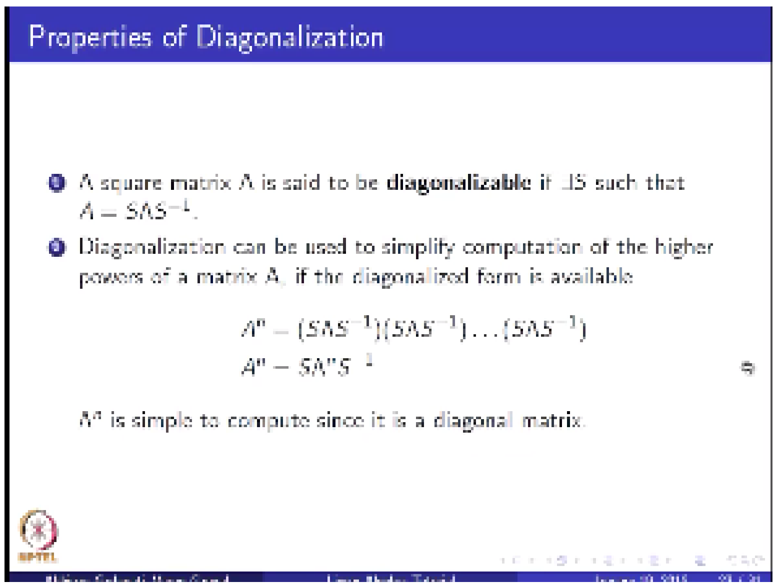


In terms of its Eigen values and eigenvectors let us consider of n cross n square matrix a we denote the matrix where every column is an eigenvector of a by s on multiplying by a each column would get multiplied by λ height since the column itself is an eigenvector of A.

This right hand side can then be simplified as the product of two matrices the first one being s itself while the second one being the diagonal matrix where the is diagonal element is the Eigen value I remember that the elitists is as now we have the equation a s equals capital laptop where capital λ is the diagonal matrix of eigenvalues on simplifying this we get e equals capital λ s inverse this is adiagonalization of a note that s inverse s is a diagonal matrix since s inverse is nothing but capital λ the diagonal matrix of Eigen values.

This result is dependent on s beinginvertibleit will hold if the eigenvalues of matrix are distinct since the eigenvectors would then be linearly independent this would mean the columns of s would be linearly independent and hence s would be full ranked and as a consequence invertible.

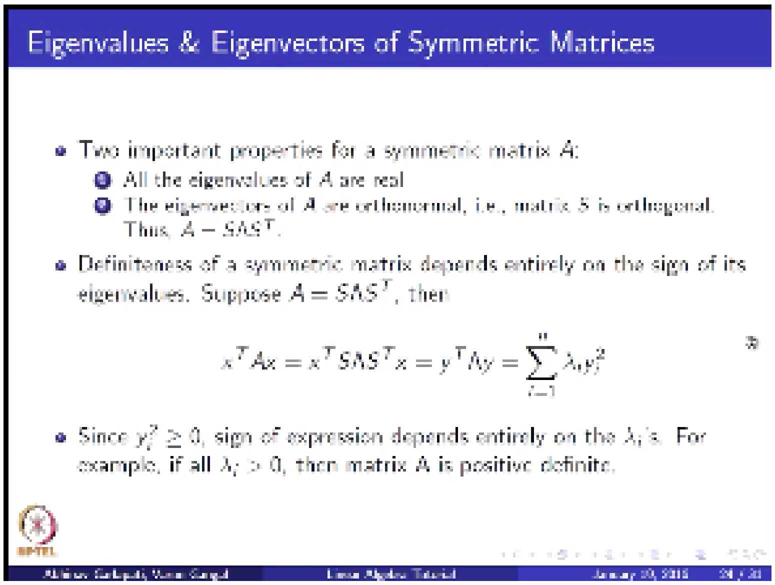
(Refer Slide Time: 08:58)



Then do we say that the square matrix isdiagonalizablewell when such a diagonalization exists we saw that we needed is to be invertible for the diagonalization to exist another advantage ofdiagonalization is that it simplifies the process of computing paper.

In we first represent every a in diagonalizedform now we can see that the s inverse of the first term and the s of the second term would multiply to give us a similarly for the second third floor and so on in this way by regrouping the terms we get a power in equal to s capital λ bar n sin verse note that it is very easy to compute the nth power of a diagonal matrix since you just have to raise every diagonal element to the power of nine this way the diagonalization has helped us simplify the process of computing a for in without this simplification. We would have needed to multiply non diagonal matrix n times if a Beatrix is symmetric.

(Refer Slide Time: 10:31)



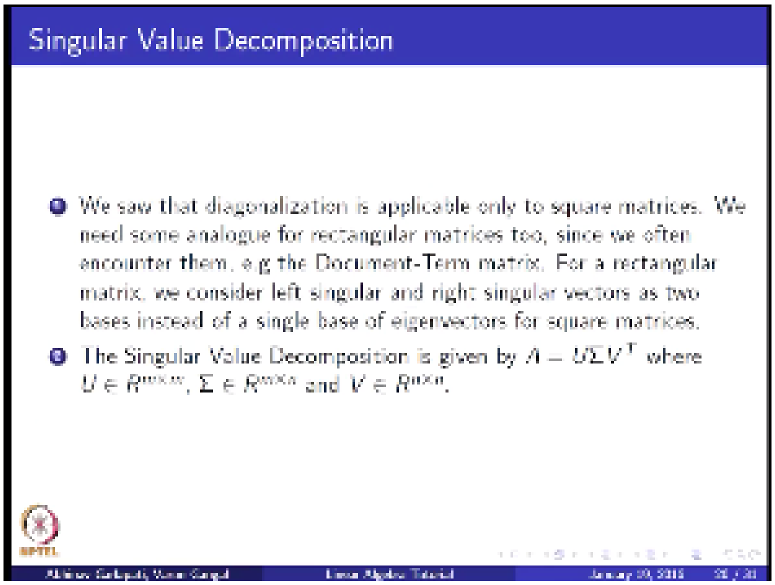
Then all its Eigen values are real numbers also its eigenvectors are also normal that is they are mutually orthogonal and normalized this means that the matrix of eigenvectors s is also orthogonal we have seen that for orthogonal matrices the inverse and the transpose are the same hence we can write a equals capital λ s transpose as furtherdiagonalization we defined our for symmetric matrices.

That definite Insane inferred from the sides of their eigenvalues suppose that a equals λ s transpose now taking the quadratic form with respect to K for vector face transpose a X simplifies to Transpose capital λ Y where Y is transpose X this further simplifies to sum over I λ I by I square now for a matrix to be positive definite this term must always be positive since Y I square is always greater than 0anyway the sign of this term depends on the Eigen values.

If all the eigenvaluesare positive the matrix is positive definite if we know that the matrix is positive semi definite or P is d then what can we say about its Eigen values since the quadratic form of a PSD matrixes non-negative for any vector X this should hold for the eigenvectors to now since ax equal to λ X transpose a X simplifies to λ norm of X square greater than equal to 0 .

Since eigenvectors are nonzero by definition the square of the norm is always positive this means that every Eigen value of axis non-negative.

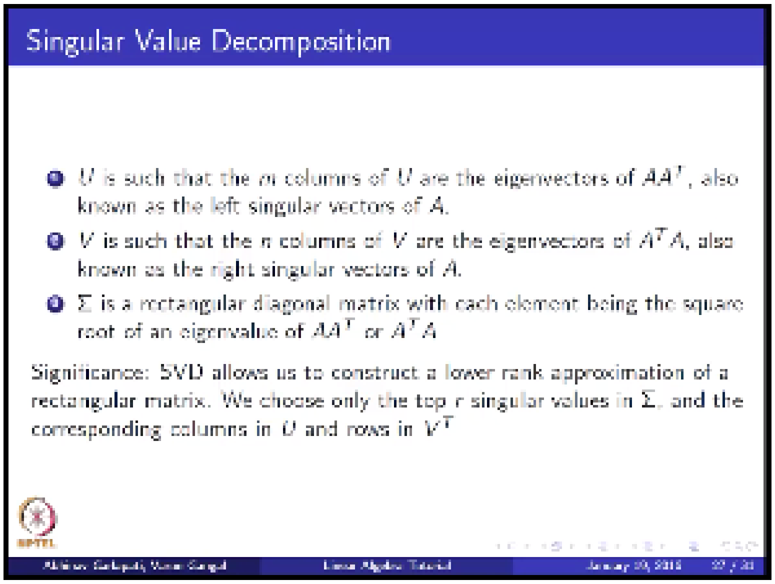
(Refer Slide Time: 13:04)



We learnt about diagonalization which took in a square matrix of size n cross n and represented it in terms of its eigenvectors however we cannot directly apply the same bygorillafor rectangular matrices since the notion of Eigen vector is defined only for a square matrix we need an I realization for rectangular matrices since we come to them often for instance the matrix of n data points or Features or the matrix of n documents and our terms for the rectangular matrix of size M cross.

In they can be predicted in terms of the Eigen vectors of a transpose and a transpose a both of which are square matrices this is known as the singular value decomposition he is represented as u Sigma V transpose where u is an M cross M matrix Sigma is an M cross n matrix and P is an N cross matrix.

(Refer Slide Time: 14:22)



The three elements u Sigma and V are as follows he knew every column represents an eigenvector of a a transpose in V every column represents an eigenvector of a transpose a Sigma is a rectangular diagonal matrix with each element being the square off an Eigen value of a a transpose or a transpose a now look at a a transpose and a transpose a have different eigenvectors but the set of Eigen values is the same this is because suppose a transpose a X equals λ X for some vacant vector X and Eigen value laughter now multiplying both sides by a we get a a transpose times ax equals λ ax hence a x is an eigenvector of a transpose Y λ is also an Eigen value of a transpose.

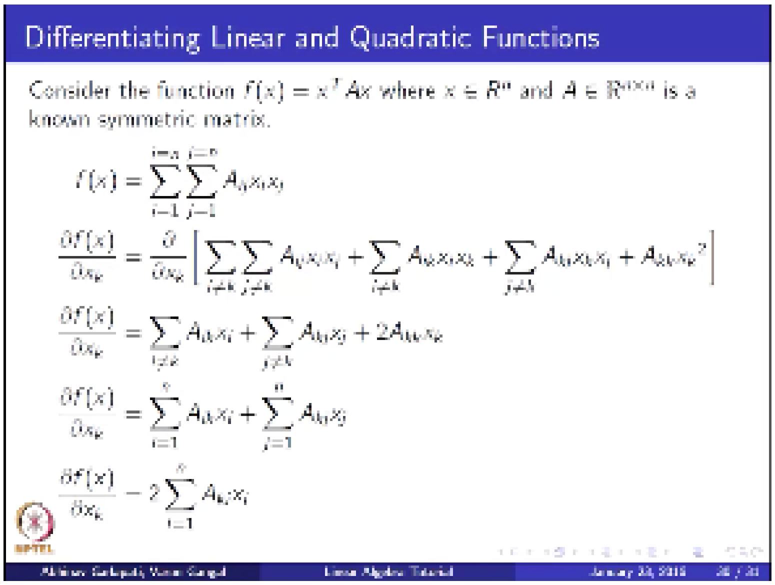
This is why a transpose they have the same set of Eigen values the significance of this decomposition is that if we ordered a human being and sigma since that the hike in values whose magnitude is large will come first goods in U and V in the column order also along the diagonal in Sigma then we can drop everything greater than he takes us to get a higher dimension and lore and approximation of the original matrix a this approximate form of a will be represented as human which is an Cross R matrix Sigma which is a R cross R matrix and V which is a n cross R matrix.

Consider a function f which takes inmate response of dimension M cross and outputs real numbers the gradient is the matrix of partial derivatives then I comma G element of Delta F of a or the gradient of f of a is the partial derivative of F of a with respect to air consider it in full time hood function which takes in a n dimensional vector and returns a real novel the Hessian for this function is defined as follows the I comma J the element of the initial is given by first differentiating f of X with respect to the J component of X HJ and in the highest component X I you can see that admission would be an N cross n matrix.

Now let us study how we can find the gradient for some simple vector functions consider the function f of X equals P transpose X where X is an Dimensional vector and B is also an n-dimensional vector off X can be written down as sum over I equals one to I equals NP IX I undifferentiating this with respect to the eighth component of the vector X we can do n things by do XK equals B K the gradient of F X is given by the vector Be can see how this intuitively relates to the first derivative of the scalar function FX equals ax which is equal to A.

We had earlier looked at a type of function called the quadratic form defined for an in cross n matrix a the quadratic form with respect to matrix axis a function f of x equals x transpose ax which takes in an n-dimensional vector X now let's have a look at how one can find the gradient and Hessian on the quadratic form of a known symmetric matrix a they can write down f of X assume over I equals 1 to n sum over Equals 1 to N K IJ X I XJ we can split-up this summation into four terms based on whether I and J are equal or not equal to K finally.

(Refer Slide Time: 18:33)



We get no effects bayou XK equal to twice somewhere I equals1 to I equals in a k i X I don't anticipation from the second last step to the last step can only be done if axis symmetric thus we get the gradient of X transpose ax is equal to 2 matrix similarly on further different changing every element of the greatly by XK we can derive the Hessian of the function the Hessian of this function comes out to be 2

IIT Madras Production

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

www.nptel.ac.in