

NPTEL
NPTEL ONLINE CERTIFICATION COURSE

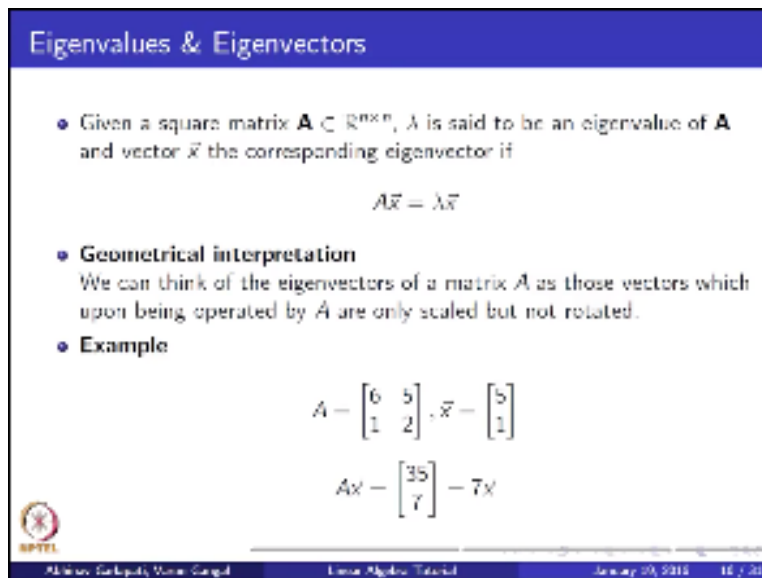
Introduction to Machine Learning

Lecture 6

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

Statistical Decision Theory –
Classification

(Refer Slide Time: 00:15)



Eigenvalues & Eigenvectors

- Given a square matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$, λ is said to be an eigenvalue of \mathbf{A} and vector \vec{x} the corresponding eigenvector if
$$\mathbf{A}\vec{x} = \lambda\vec{x}$$
- Geometrical interpretation**
We can think of the eigenvectors of a matrix \mathbf{A} as those vectors which upon being operated by \mathbf{A} are only scaled but not rotated.
- Example**
$$\mathbf{A} = \begin{bmatrix} 6 & 5 \\ 1 & 2 \end{bmatrix}, \vec{x} = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$$
$$\mathbf{A}\vec{x} = \begin{bmatrix} 35 \\ 7 \end{bmatrix} = 7\vec{x}$$

NPTEL
Ankur, Gurbir, Vignesh, Chandan
Linear Algebra Tutorial
January 09, 2018 18 / 23

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)

Characteristic Equation

- Trivially, the $\vec{0}$ vector would always be an eigenvector of any matrix. Hence, we only refer only to non-zero vectors as eigenvectors.
- Given a matrix A , how do we find all eigenvalue-eigenvector pairs?

$$A\vec{x} = \lambda\vec{x}$$

$$A\vec{x} - \lambda\vec{x} = \vec{0}$$

$$(A - \lambda I)\vec{x} = \vec{0}$$

The above will hold iff

$$|(A - \lambda I)| = 0$$

This equation is also referred to as the characteristic equation of A . Solving the equation gives us all the eigenvalues λ of A . Note that these eigenvalues can be **complex**.

MIT OpenCourseWare
Linear Algebra, Gilbert Strang
January 29, 2015 17 / 21

We can see that you know would always be an eye effectors of any matrix if we simply go by the ax equals λX 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 λX we get a $-\lambda I$ $\vec{x} = \vec{0}$ now since we are only looking at nonzero vectors in no fix cannot be 0 and expand be a zero vector which means that picked up a - laughter should pay zero so the equation that of a $-\lambda$ 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)

Properties

- The trace $\text{tr}(A)$ of a matrix A also equals the sum of its n eigenvalues.

$$\text{tr}(A) = \sum_{i=1}^n \lambda_i$$

- The determinant $|A|$ is equal to the product of the eigenvalues.

$$|A| = \prod_{i=1}^n \lambda_i$$

- The rank of a matrix is equal to the number of non zero eigenvalues of A .
- If A is invertible, then the eigenvalues of A^{-1} are of form $\frac{1}{\lambda_i}$, where λ_i are the eigenvalues of A .

NPTEL
 Anil Kumar, Debajyoti, Vimal, Ganguli
 Linear Algebra, Tamil Nadu
 January 09, 2015 18 / 28

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 x_1 and x_2 both 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 $\frac{1}{\lambda}$ where λ is an Eigen value of A .

(Refer Slide Time: 03:55)

Proof

$$\sum_{i=1}^{i=k} a_i v_i = \vec{0}$$

$$(A - \lambda_k I) \sum_{i=1}^{i=k} a_i v_i = \vec{0}$$

$$\sum_{i=1}^{i=k} (A - \lambda_k I) a_i v_i = \vec{0} \quad \circledast$$

$$\sum_{i=1}^{i=k} a_i (\lambda_i - \lambda_k) v_i = \vec{0}$$

Since the eigenvalues are distinct, $\lambda_i \neq \lambda_k \forall i \neq k$. Thus the set of $(k - 1)$ eigenvectors is also linearly dependent, violating our assumption of it being the smallest such set. This is a result of our incorrect starting assumption. Hence proved by contradiction.

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 eigenvalues distinct in its Eigen vectors are linearly independent we shall prove this by what is called a proof by contradiction if 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 $\sum_{i=1}^k a_i v_i = \vec{0}$ now let us multiply both sides of the equation by $(A - \lambda_k I)$ since v_k is an eigenvector of $(A - \lambda_k I)$ will be equal to zero we can understand this from the characteristic equation hence the term corresponding to v_k disappears from the equation since it goes to zero.

Now for the remaining Eigen values since we know they are distinct the $(A - \lambda_k I)$ cannot be equal to zero note that $(A - \lambda_k I) \sum_{i=1}^k a_i v_i = \vec{0}$ simplifies to $\sum_{i=1}^k a_i (\lambda_i - \lambda_k) v_i = \vec{0}$ since $\sum_{i=1}^k a_i v_i = \vec{0}$ we can think of $\sum_{i=1}^k a_i (\lambda_i - \lambda_k) v_i = \vec{0}$ as a new constant P this means now that we have a summation running from $i=1$ to $i=k-1$ such that $\sum_{i=1}^{i=k-1} a_i v_i = \vec{0}$ 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 2 hence 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)

Diagonalization

Given a matrix A , we consider the matrix S with each column being an eigenvector of A

$$S = \begin{bmatrix} | & | & \dots & | \\ v_1 & v_2 & \dots & v_n \\ | & | & \dots & | \end{bmatrix}$$

$$AS = \begin{bmatrix} | & | & \dots & | \\ \lambda_1 v_1 & \lambda_2 v_2 & \dots & \lambda_n v_n \\ | & | & \dots & | \end{bmatrix}$$

$$AS = \begin{bmatrix} | & | & \dots & | \\ v_1 & v_2 & \dots & v_n \\ | & | & \dots & | \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \dots \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n \end{bmatrix}$$


© 2014 NPTEL. All rights reserved. Linear Algebra, Tutorial 1, January 09, 2014, 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 hence 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 diagonal element is the Eigen value. I remember that the equation is as now we have the equation $AS = S\Lambda$ where Λ is the diagonal matrix of eigenvalues on simplifying this we get $S^{-1}AS = \Lambda$ inverse this is diagonalization of A note that $S^{-1}S$ is a diagonal matrix since $S^{-1}S$ is nothing but I the diagonal matrix of Eigen values.

This result is dependent on S being invertible. It will hold if the eigenvalues of matrix A 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)

Properties of Diagonalization

- 1 A square matrix A is said to be **diagonalizable** if $\exists S$ such that $A = SAS^{-1}$.
- 2 Diagonalization can be used to simplify computation of the higher powers of a matrix A , if the diagonalized form is available

$$A^n = (SAS^{-1})(SAS^{-1}) \dots (SAS^{-1})$$

$$A^n = SA^nS^{-1}$$

A^n is simple to compute since it is a diagonal matrix.

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

In we first represent every A in diagonalized form now we can see that the S^{-1} 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 $S^{-1} \Lambda^n S$ inverse note that it is very easy to compute the n th power of a diagonal matrix since you just have to raise every diagonal element to the power of n this way the diagonalization has helped us simplify the process of computing A^n for in without this simplification. We would have needed to multiply non diagonal matrix n times if a matrix is symmetric.

(Refer Slide Time: 10:31)

Eigenvalues & Eigenvectors of Symmetric Matrices

- Two important properties for a symmetric matrix A :
 - All the eigenvalues of A are real
 - The eigenvectors of A are orthonormal, i.e., matrix S is orthogonal. Thus, $A = SAS^T$.
- Definiteness of a symmetric matrix depends entirely on the sign of its eigenvalues. Suppose $A = SAS^T$, then

$$x^T Ax = x^T SAS^T x = y^T \Lambda y = \sum_{i=1}^n \lambda_i y_i^2 \quad \text{Ⓜ}$$

- Since $y_i^2 \geq 0$, sign of expression depends entirely on the λ_i 's. For example, if all $\lambda_i > 0$, then matrix A is positive definite.

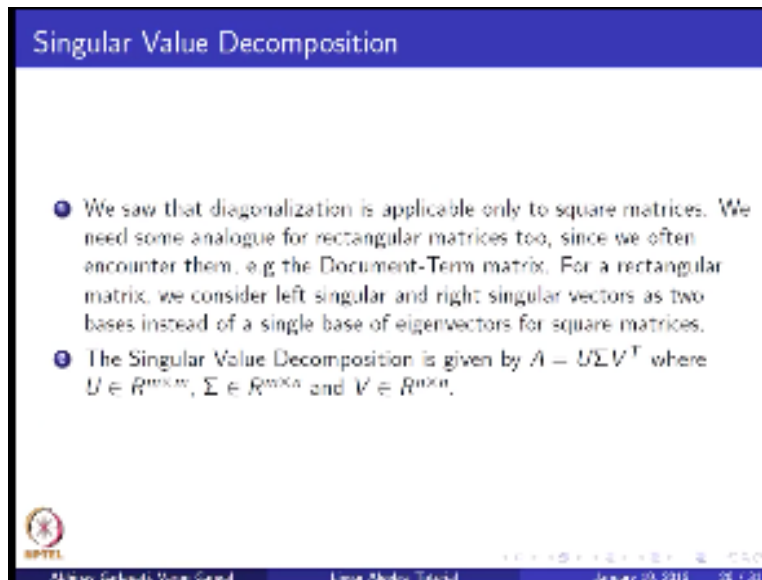


Then all its Eigen values are real numbers also its eigenvectors are also normal that is they are mutually orthogonal and normalized this means that the matrix of eigenvectors s is also orthogonal we have seen that for orthogonal matrices the inverse and the transpose are the same hence we can write $A = S \Lambda S^T$ as further diagonalization we defined our for symmetric matrices.

That definite is inferred from the signs of their eigenvalues suppose that $A = S \Lambda S^T$ now taking the quadratic form with respect to x for vector x transpose $x^T A x$ simplifies to $y^T \Lambda y$ where y is transpose x this further simplifies to sum over i $\lambda_i y_i^2$ now for a matrix to be positive definite this term must always be positive since y_i^2 is always greater than 0 anyway the sign of this term depends on the Eigen values.

If all the eigenvalues are positive the matrix is positive definite if we know that the matrix is positive semi definite or PSD then what can we say about its Eigen values since the quadratic form of a PSD matrix is non-negative for any vector x this should hold for the eigenvectors to now since $x^T A x = \lambda x^T x$ simplifies to $\lambda \|x\|^2$ 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 PSD matrix is non-negative.

(Refer Slide Time: 13:04)



Singular Value Decomposition

- 1 We saw that diagonalization is applicable only to square matrices. We need some analogue for rectangular matrices too, since we often encounter them, e.g. the Document-Term matrix. For a rectangular matrix, we consider left singular and right singular vectors as two bases instead of a single base of eigenvectors for square matrices.
- 2 The Singular Value Decomposition is given by $A = U\Sigma V^T$ where $U \in R^{m \times m}$, $\Sigma \in R^{m \times n}$ and $V \in R^{n \times n}$.

NPTEL
Ankur Gupta, Wipro, Google | Linear Algebra Tutorial | January 09, 2018 | 05 / 20

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)

Singular Value Decomposition

- ① U is such that the m columns of U are the eigenvectors of AA^T , also known as the left singular vectors of A .
- ② V is such that the n columns of V are the eigenvectors of $A^T A$, also known as the right singular vectors of A .
- ③ Σ is a rectangular diagonal matrix with each element being the square root of an eigenvalue of AA^T or $A^T A$.

Significance: SVD allows us to construct a lower rank approximation of a rectangular matrix. We choose only the top r singular values in Σ , and the corresponding columns in U and rows in V^T .


© Anirban Ghoshal, IITK, Mumbai
Linear Algebra, Tutorial 4
January 09, 2015 07 / 20

The three elements U , Σ and V are as follows. Every column represents an eigenvector of a transpose in V . Every column represents an eigenvector of a transpose. Σ is a rectangular diagonal matrix with each element being the square root of an Eigen value of a transpose or a transpose. Now look at a transpose and a transpose. They have different eigenvectors but the set of Eigen values is the same. This is because suppose a transpose A equals λX for some vector X and Eigen value λ . Now multiplying both sides by A^T we get $A^T A X = \lambda A^T X$. Hence X is an eigenvector of a transpose $A^T A$. 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 Σ 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 Σ . Then we can drop everything greater than he takes us to get a higher dimension and low rank approximation of the original matrix. This approximate form of A will be represented as $U_r \Sigma_r V_r^T$ where U_r is an m cross r matrix, Σ_r which is a r cross r matrix and V_r which is a n cross r matrix.

Consider a function f which takes an input vector of dimension n and outputs real numbers. The gradient is the matrix of partial derivatives. The i -th component of the gradient of f at \mathbf{x} is the partial derivative of f at \mathbf{x} with respect to x_i . Consider it in full time. A Hessian function which takes in an n -dimensional vector and returns a real number. The Hessian for this function is defined as follows: the i, j -th element of the Hessian is given by first differentiating f of \mathbf{x} with respect to the j -th component of \mathbf{x} , and then differentiating that result with respect to the i -th component of \mathbf{x} . You can see that the Hessian would be an $n \times n$ matrix.

Now let us study how we can find the gradient for some simple vector functions. Consider the function $f(\mathbf{x}) = \mathbf{b}^T \mathbf{x}$ where \mathbf{x} is an n -dimensional vector and \mathbf{b} is also an n -dimensional vector. $f(\mathbf{x})$ can be written down as $\sum_{i=1}^n b_i x_i$. Differentiating this with respect to the i -th component of the vector \mathbf{x} , we can do n things by doing $\frac{\partial f}{\partial x_i} = b_i$. The gradient of $f(\mathbf{x})$ is given by the vector \mathbf{b} . We can see how this intuitively relates to the first derivative of the scalar function $f(x) = ax$ which is equal to a .

We had earlier looked at a type of function called the quadratic form defined for an $n \times n$ matrix \mathbf{A} . The quadratic form with respect to matrix \mathbf{A} is a function $f(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x}$ which takes in an n -dimensional vector \mathbf{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 \mathbf{A} . They can write down $f(\mathbf{x}) = \sum_{i=1}^n \sum_{j=1}^n A_{ij} x_i x_j$. 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)

Differentiating Linear and Quadratic Functions

Consider the function $f(x) = x^T A x$ where $x \in \mathbb{R}^n$ and $A \in \mathbb{R}^{n \times n}$ is a known symmetric matrix.

$$f(x) = \sum_{i=1}^n \sum_{j=1}^n A_{ij} x_i x_j$$

$$\frac{\partial f(x)}{\partial x_k} = \frac{\partial}{\partial x_k} \left[\sum_{i \neq k} \sum_{j \neq k} A_{ij} x_i x_j + \sum_{i \neq k} A_{ik} x_i x_k + \sum_{j \neq k} A_{kj} x_k x_j + A_{kk} x_k^2 \right]$$

$$\frac{\partial f(x)}{\partial x_k} = \sum_{i \neq k} A_{ik} x_i + \sum_{j \neq k} A_{kj} x_j + 2A_{kk} x_k$$

$$\frac{\partial f(x)}{\partial x_k} = \sum_{i=1}^n A_{ik} x_i + \sum_{j=1}^n A_{kj} x_j$$

$$\frac{\partial f(x)}{\partial x_k} = 2 \sum_{i=1}^n A_{ki} x_i$$



NPTEL

Linear Algebra, Tamil

January 23, 2015

26 / 31

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

Copyrights Reserved