

**NPTEL**

**NPTEL ONLINE CERTIFICATION COURSE**

**Linear Algebra Tutorial  
CS5011 – Machine Learning**

**Abhinav Garlapati Varun Gangal**

**Department of Computer Science  
IIT Madras**

**January 17, 2016**

Hi everyone welcome to the second tutorial of this of the introduction to machine learning course, so in this tutorial we shall be taking a tour of the aspects of linear algebra which you would need for the course we will cover a variety of concepts and subspaces basis span and become a decompositions eigenvalues, eigenvectors over the course of the tutorial.

(Refer Slide Time: 00:47)

**What is Linear Algebra**

**Linear Algebra**

Linear algebra is the branch of mathematics concerning vector spaces and linear mappings between such spaces. It includes the study of lines, planes, and subspaces, but is also concerned with properties common to all vector spaces.

Why do we study Linear Algebra?

- Provides a way to compactly represent & operate on sets of linear equations.
- In machine learning, we represent data as matrices and hence it is natural to use notions and formalisms developed in Linear Algebra.

NPTEL  
Abhinav Garlapati, Varun Gangal      Linear Algebra Tutorial      January 17, 2016      2 / 31

So the first question one would ask is, why we need linear algebra at all and what is linear algebra? So you may have come across this in school or your +1 or +2 level but just to recap I mean linear algebra is the branch of mathematics which deals with vectors and make the big vector spaces and linear mappings between these spaces so, why do we study linear algebra here

and especially the context of machine learning is firstly it gives us a way to manipulate represent and operate sets of linear equations but why do these linear equations pop up in machine learning in the first place, so the reason for that is not machine learning we are represent our data as an  $n \times p$  matrix open where  $n$  is the number of data points and  $P$  is the number of features so it is natural we have to use motions and formalisms developed in linear algebra.

What is the data even I mean the parameter is you use are represented as vectors so as a result linear algebra has an important role to play in machine learning.

(Refer Slide Time: 02:05)

The slide is titled "Introduction to LinAl" and contains the following content:

- Consider the following system of equations:
$$\begin{aligned}4x_1 - 5x_2 &= -13 \\ -2x_1 + 3x_2 &= 9\end{aligned}$$
- In matrix notation, the system is more compactly represented as:
$$Ax = b$$
$$A = \begin{bmatrix} 4 & -5 \\ -2 & 3 \end{bmatrix}$$
$$b = \begin{bmatrix} -13 \\ 9 \end{bmatrix}$$

At the bottom of the slide, there is a logo for "MPTL" and a footer with the text "Abhinav Garlapati, Varun Gangal", "Linear Algebra Tutorial", "January 17, 2016", and "3 / 31".

So you can see where a system of linear equations two equations and two variables  $4x_1 - 5x_2 = -13$  and  $-2x_1 + 3x_2 = 9$ , so we can right away see here that advantages of a matrix notation, so if you see below you can represent the same system of two equations directly as one equation in the form of  $ax = b$  where  $a$  is the set of coefficients and  $X$  is the  $2 \times 1$  matrix two rows and one column or you may also call it a 2-dimensional vector  $X \ 1 \times 2$ .

So we can see that when you multiply the matrix  $A$  with  $X \ 1 \times 2$  the  $2 \times 1$  matrix you will get back the same L.H.S or the set of two L.H.S and  $P$  the matrix  $B$  represents the RHS so we can it is very easy to verify that if you represent this in matrices you can get you will get back the original representation and from that representation you can get this so to solving this in general without matrices how we would solve this is you will first solve for one variable and then

substitute to get the other but this using matrices you can even solve this directly, so just multiply both the sides by a inverse.

So you would get  $X$  and  $A^{-1}B$  of course when you want you would have to care about is that all matrices do not have an inverse but if but in most cases they do, so in that case you can directly get the solution of  $X$  in the form of  $x = A^{-1}B$  were solving it just to one equation so as we said earlier I mean linear algebra gives us this freedom to manipulate several equations at once and multiple variables.

(Refer Slide Time: 04:10)

**Vector Space**

**Definition**  
A set  $V$  with two operations  $+$  and  $\cdot$  is said to be a **vector space** if it is closed under both these operations and satisfies the following eight axioms.

- 1 Commutative Law  
$$x + y = y + x, \quad \forall x, y \in V$$
- 2 Associative Law  
$$(x + y) + z = x + (y + z), \quad \forall x, y, z \in V$$
- 3 Additive identity  
$$\exists 0 \in V \text{ s.t. } x + 0 = x, \quad \forall x \in V$$
- 4 Additive inverse  
$$\forall x \in V, \exists \bar{x} \text{ s.t. } x + \bar{x} = 0$$

MPTEL  
Abhinav Garlapati, Varun Gangal      Linear Algebra Tutorial      January 19, 2016      4 / 31

A fundamental definition in linear algebra is that of a vector space a set of vectors  $V$  is said to be a vector space if it is closed under the operations of vector addition and scalar multiplication and in addition satisfies the axioms we have distributed, like knows to be mean that if we take two elements from this set  $X$  &  $Y$  then  $X + Y$  will also lies in the sector  $V$ , in addition to this if we take her in scalar  $\alpha$  a real number that is and multiply a vector from this set by with it then  $\alpha y$  also belongs to  $v$ , in both these properties are satisfied then the set of vectors is said to be closed with respect to vector addition and scalar multiplication, now let us have a look at the axioms. The first one is a commutative law, the commutative law states that if you pick any two elements from the set  $v$ ,  $x$  and  $y$  then  $x + y = y + x$ , the associative law says that if you pick any point from this set  $X$   $Z$  then  $X + Y$  added together  $+ Z = X + Y + Z$  added together.

The additive identity loss is that there exists an additive identity or the 0 so as to say in the set since that if you pick any element from the set and add this 0 to it you get back the same element, the additive inverse losses is that there exists for every element there exists another for every element X there exists a corresponding  $\bar{x}$  such then  $X + \bar{x} = 0$ .

(Refer Slide Time: 06:16)

**Vector Space (Contd.)**

- 1 Distributive Law  

$$\alpha \cdot (x + y) = \alpha \cdot x + \alpha \cdot y, \quad \forall \alpha \in \mathbb{R}, x, y \in V$$
- 2 Distributive Law  

$$(\alpha + \beta) \cdot x = \alpha \cdot x + \beta \cdot x, \quad \forall \alpha, \beta \in \mathbb{R}, x \in V$$
- 3 Associative Law  

$$(\alpha\beta) \cdot x = \alpha \cdot (\beta \cdot x), \quad \forall \alpha, \beta \in \mathbb{R}, x \in V$$
- 4 Unitary Law  

$$1 \cdot x = x, \quad \forall x \in V$$

NPTEL  
 Abhinav Garlapati, Varun Gangal      Linear Algebra Tutorial      January 19, 2016      5 / 31


The fifth law is the distributive law, this law says that, if you have a real scalar  $\alpha$  with which you multiply the sum of two vectors  $X + Y$  then that should be equal to  $\alpha x + \alpha y$  the second distributive law says that if you have the sum of two scalars multiplying a vector X from the same  $\alpha + \beta ( a. x)$  then that should be equal to  $\alpha$  times X +  $\beta$  times X, if associative law says that if you will first multiply two scalars and then multiplying the vector with them that should be equal to multiplying it first with the second scalar  $\beta$  and then with  $\alpha$ . The unitary law says that on multiplication by the scalar real number one you get back the same vector so this is important because you would want multiplication to force any unexpected scaling like if you multiplied by the scalar K is a vector you can surely be exactly K times it should not be say rote kit.

(Refer Slide Time: 07:35)

## Subspace

**Definition**  
Let  $W$  be a subset of a vector space  $V$ . Then  $W$  is called a **subspace** of  $V$  if  $W$  is a vector space.

- Do we have to verify all 8 conditions to check whether a given subset of a vector space is a subspace?
- **Theorem:** Let  $W$  be a subset of a vector space  $V$ . Then  $W$  is a subspace of  $V$  if and only if  $W$  is non-empty and  $x + \alpha y \in W, \forall x, y \in W, \alpha \in \mathbb{R}$



Abhinav Garlapati, Varun Gangal      Linear Algebra Tutorial      January 19, 2016      6 / 31

A second related definition is that of a subspace, a subset  $W$  of a vector space  $V$  is said to be a subspace if  $W$  is a vector space now this means that  $W$  should be closed under vector addition and scalar multiplication it should also satisfy the eight axioms we stated earlier, now the question that arises is to begin to verify all these eight conditions given that we know that  $W$  is already a subset of a vector space it is enough to check just for the following two conditions first means that  $W$  is non empty that in other words it has at least a single element.

And secondly that if I pick any two elements  $x$  and  $y$  from this set and any real number  $\alpha$  then  $x + \alpha y$  should belong to  $W$ .

(Refer Slide Time: 08:39)

**Norm**

**Definition**


Norm is any function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  satisfying:

- 1  $\forall x \in \mathbb{R}^n, f(x) \geq 0$  (non-negativity)
- 2  $f(x) = 0$  iff  $x = 0$  (definiteness)
- 3  $\forall x \in \mathbb{R}^n, f(tx) = |t|f(x)$  (homogeneity)
- 4  $\forall x, y \in \mathbb{R}^n, f(x+y) \leq f(x) + f(y)$  (triangle inequality)

- Example -  $l_p$  norm

$$\|x\|_p = \left( \sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}}$$

- Matrices can have norms too - e.g., Frobenius norm

$$\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n A_{ij}^2} = \sqrt{\text{tr}(A^T A)} \quad (1)$$


Abhinav Garlapati, Varun Gargal      Linear Algebra Tutorial      January 19, 2016      7 / 31

Now let us have a look at the definition of a norm, so intuitively a norm is a measure of the length of a vector or its magnitude it is a function from this vector space which mostly happens to be  $\mathbb{R}^n$  where  $n$  is a dimension of a vector to the space of real numbers  $\mathbb{R}$ , so for a function to be a norm it should satisfy the four conditions which we have given here, firstly it should be it should always be non-negative secondly it should be zero if and only if the vector is zero.

Thirdly for every for every vector if you multiplied by a scalar its norm should get multiplied by the body modulus of the scalar by modulus here we mean the absolute value and both being that if we take any pair of vectors in our vector space which is  $\mathbb{R}^n$  the norm of the sum of these two vectors should be less than some of their norms which this is also known as straggle inequality. So this is namely related to the fact that the third side of attractor should always be less than some of the other two sides.

Now an example of a norm is the LP norm then you sum up the absolute values along each dimension raised to P and then take the 1/P as root of this, so when  $P = 2$  you get the A2 norm which is the magnitude of a vector as we have learnt in our earlier studies  $\sqrt{x^2 + y^2}$  if you are looking at just the space  $\mathbb{R}^2$  there are other norms for instance there are norms defined for matrices as well here we had totally defined all for effect for vectors so the Frobenius norm is a matrix norm so what it does is it essentially sums of the squares of all the elements and then takes the root of that, so this also happens to be equal to the trace of  $A^{-1} A$  now the trace of a matrix is simply the some of its diagonal elements.

(Refer Slide Time: 11:12)

**Range Of A Matrix**

- The **span** of a set of vectors  $X = \{x_1, x_2, \dots, x_n\}$  is the set of all vectors that can be expressed as a linear combination of the vectors in  $X$ .  
In other words, set of all vectors  $v$  such that  $v = \sum_{i=1}^{|X|} \alpha_i x_i, \alpha_i \in R$
- The **range** or **columnspace** of a matrix  $A$ , denoted by  $R(A)$  is the span of its columns. In other words, it contains all linear combinations of the columns of  $A$ . For instance, the column space of  $A = \begin{bmatrix} 1 & 0 \\ 5 & 4 \\ 2 & 4 \end{bmatrix}$  is the plane spanned by the vectors  $\begin{bmatrix} 1 \\ 5 \\ 2 \end{bmatrix}$  and  $\begin{bmatrix} 0 \\ 4 \\ 4 \end{bmatrix}$ .

MPTCL  
Abhinav Garlapati, Varun Gangal      Linear Algebra Tutorial      January 19, 2016      8 / 31

This span of a set of vectors is the set of all vectors which can be composed using these vectors by using the operations of vector addition and scalar multiplication, so the name span comes from the fact that this set of vectors spans a potentially larger set of vectors which is then called a span. So to define this more formally it is the set of all vectors  $V$  such that  $V = \sum_{i=1}^{|X|} \alpha_i x_i$  where  $\alpha_i$  is a real number.

Now a related definition is that of range or column space, so if we think of a matrix each of its columns is a vector so the set of all columns of a matrix is like a set of vectors now the span of this set of vectors is called the range or column space of that matrix, so if you consider the matrix given here the columns of this matrix  $A$  are  $\begin{bmatrix} 1 \\ 5 \\ 2 \end{bmatrix}$  and  $\begin{bmatrix} 0 \\ 4 \\ 4 \end{bmatrix}$ , so what would be this column space of this matrix, it would be this the span of the vectors  $\begin{bmatrix} 1 \\ 5 \\ 2 \end{bmatrix}$  and  $\begin{bmatrix} 0 \\ 4 \\ 4 \end{bmatrix}$  which is essentially the plane which is spanned by these two vectors.

(Refer Slide Time: 12:56)


**Nullspace Of A Matrix**

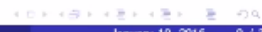
**Definition**

The nullspace  $N(A)$  of a matrix  $A \in \mathbb{R}^{m \times n}$  is the set of all vectors that equal 0 when multiplied by  $A$ . The dimensionality of the nullspace is also referred to as the **nullity** of  $A$ .

$$N(A) = \{x \in \mathbb{R}^n : Ax = 0\}$$

- Note that vectors in  $N(A)$  are of dimension  $n$ , while those in  $R(A)$  are of size  $m$ , so vectors in  $R(A^T)$  and  $N(A)$  are both of dimension  $n$ .





Abhinav Garlapati, Varun Gargal
Linear Algebra Tutorial
January 19, 2016 9 / 31

If we have matrix  $A$  of dimensions  $m \times n$  then the null space is the set of  $n \times 1$  vectors which given  $m \times 1$  0 vector on  $B \times A$ , in other words it is the set of vectors  $X$  such that  $ax = 0$  nullity is the rank or dimensionality of the null space we will be within the definition of nullity later once you defined rank more clearly, another interesting fact about null spaces is that the null space of  $A$  and not  $B$  is of dimension  $n$  while the range of  $A$  or the column space as we defined it earlier is of dimension  $M$  or  $M \times 1$ .

This means that vectors in  $R(A^{-1})$  so note that  $A^{-1}$  is  $n \times M$  So vector in the range of  $A^{-1}$  will be of dimension  $n$  similar to the vectors in the null space of  $A$ , so this means that the vectors in range of  $A^{-1}$  and null space of  $A$  would both be of the dimension  $A \times 1$ .

(Refer Slide Time: 14:15)



## Example

Consider the matrix

$$A = \begin{bmatrix} 1 & 0 \\ 5 & 4 \\ 2 & 4 \end{bmatrix}$$

The nullspace of  $A$  is made up of vectors  $x$  of the form  $\begin{bmatrix} u \\ v \end{bmatrix}$ , such that

$$\begin{bmatrix} 1 & 0 \\ 5 & 4 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

The nullspace here only contains the vector  $(0,0)$ .



Let us consider an example to illustrate the concept of a null space, consider the matrix  $A$  given here,  $A$  is a  $3 \times 2$  matrix hence the null space of  $A$  will be made up of vectors of dimension  $2$  or  $2 \times 1$ . Now we see that the on solving being weakened  $u = 0$ ,  $v = 0$  this means that the null space only contains the  $0$  vector the two dimensional  $0$  vector  $0, 0$ .

(Refer Slide Time: 14:52)

## Another Example

Now, consider the matrix

$$B = \begin{bmatrix} 1 & 0 & 1 \\ 5 & 4 & 9 \\ 2 & 4 & 6 \end{bmatrix}$$

Here, the third column is a linear combination of the first two columns.  
Here, the nullspace is the line of all points  $x = c, y = c, z = -c$ .



Let us consider another example to illustrate null space is fitted, take the matrix B which is a 3 x 3 matrix the null space would consist of 3 x 1 vectors we leave the finding of the null space to the audiences and accessories however when on solving we get the null space to be the set of all vectors of the form  $x = c, y = C$  and  $z = -c$ , where C is any real number and XYZ referred to the first second and third dimension respectively.

(Refer Slide Time: 15:35)

## Linear Independence and Rank

### Definition

A set of vectors  $\{x_1, x_2, \dots, x_n\} \in \mathbb{R}^n$  is said to be **(linearly) independent** if no vector can be represented as a linear combination of the remaining vectors.

- i.e., if  $x_n = \sum_{i=1}^{n-1} \alpha_i x_i$  for some scalar values  $\alpha_1, \dots, \alpha_{n-1} \in \mathbb{R}$ , then we say that the vectors  $\{x_1, x_2, \dots, x_n\}$  are linearly dependent; otherwise, the vectors are linearly independent
- The **column rank** of a matrix  $A \in \mathbb{R}^{m \times n}$  is the size of the largest subset of columns of  $A$  that constitute a linearly independent set
- Similarly, **row rank** of a matrix is the largest number of rows of  $A$  that constitute a linearly independent set



Before we dive into defining linear independence recollect how we define the linear combination, a set of vectors is linearly independent if no vector in the set can be produced using the linear combination of the other vectors in the set. Now let us have a look at the related concept of rank, so the column rank of  $m \times n$  matrix  $A$  is the size of the largest linearly independent subset of columns. Note that our columns here are  $m \times 1$  vectors the row rank is defined in a similar way for rows.

(Refer Slide Time: 16:22)

Properties Of Ranks

- For any matrix  $A \in \mathbb{R}^{m \times n}$ , it turns out that the column rank of  $A$  is equal to the row rank of  $A$ , collectively as the rank of  $A$ , denoted as **rank(A)**
- Some basic properties of the rank:
  - 1 For  $A \in \mathbb{R}^{m \times n}$ ,  $\text{rank}(A) \leq \min(m, n)$ .  
If  $\text{rank}(A) = \min(m, n)$ ,  $A$  is said to be **full rank**
  - 2 For  $A \in \mathbb{R}^{m \times n}$ ,  $\text{rank}(A) = \text{rank}(A^T)$
  - 3 For  $A \in \mathbb{R}^{m \times n}$ ,  $B \in \mathbb{R}^{n \times p}$ ,  $\text{rank}(AB) \leq \min(\text{rank}(A), \text{rank}(B))$  ☼
  - 4 For  $A, B \in \mathbb{R}^{m \times n}$ ,  $\text{rank}(A + B) \leq \text{rank}(A) + \text{rank}(B)$

MPTEL  
Abhinav Garlapati, Varun Gangal Linear Algebra Tutorial January 19, 2016 13 / 31


Let us walk through some interesting properties of ranks, for any  $m \times n$  matrix of real numbers the column rank is equal to the row rank and we refer to this quantity as the rank of the matrix, earlier we had looked at the quantity called nullity, nullity is the rank of the null space of  $A$  some other interesting properties of ranks are also listed here, for instance the rank of a matrix is utmost the minimum of its two dimensions.

The row dimension  $\times$  column dimension secondly the rank of a matrix is the same as the rank of its transpose, thirdly if you multiply two matrices  $A$  and  $B$  the rank of the resultant matrix is utmost the minimum of the ranks of  $A$  and  $B$  if you add up two matrices the rank of the resultant matrix is utmost the sum of the ranks of  $A$  and  $B$ .

(Refer Slide Time: 17:33)

## Orthogonal Matrices

- A square matrix  $U \in \mathbb{R}^{n \times n}$  is **orthogonal** iff
  - All columns are mutually orthogonal  $v_i^T v_j = 0, \forall i \neq j$
  - All columns are normalized  $v_i^T v_i = 1, \forall i$
- If  $U$  is orthogonal,  $UU^T = U^T U = I$ . This also implies that the inverse of  $U$  happens to be its transpose.
- Another salient property of orthogonal matrices is that **they do not change** the Euclidean norm of a vector when they operate on it, i.e.  $\|Ux\|_2 = \|x\|_2$ .  
 Multiplication by an orthogonal matrix can be thought of as a pure rotation, i.e., it does not change the magnitude of the vector, but changes the direction.


Abhinav Gargapati, Varun Gangal
Linear Algebra Tutorial
January 19, 2016 14 / 31

A square matrix  $U$  of dimension  $n \times n$  is defined to be orthogonal if and only if the following two conditions would firstly all pairs of distinct columns should be orthogonal by columns being orthogonal we mean that the dot product of any pair of distinct column vectors is zero in other words  $v_i^T v_j = 0$  for all  $i \neq j$  the second condition is that the dot product on any column with itself or  $v_i^T v_i = 1$ .

In other words all the column vectors should be normalized, an interesting implication of a matrix being orthogonal is that  $UU^T$  and  $U^T U$  both end up being equal to the  $n \times n$  identity matrix  $I$  this also means that  $U^{-1} = U^T$  equals the transpose of such an orthogonal matrix is also means equals and additional interesting property is seen when we multiply a  $n \times 1$  vector  $x$  by  $n \times n$  orthogonal matrix  $U$  the Euclidean or  $L_2$  norm of such a vector  $x$  remains the same on multiplication by  $U$ .

It actively we can understand this as orthogonal matrices  $U$  performing only pure rotation on multiplying the vector  $x$  in other words they only change the direction of a vector but do not change its magnitude.

(Refer Slide Time: 19:27)

## Quadratic Form of Matrices

- Given a square matrix  $A \in \mathbb{R}^{n \times n}$  and a vector  $x \in \mathbb{R}^n$ , the scalar value  $x^T A x$  is called a **quadratic form**
- A symmetric matrix  $A \in \mathbb{S}^n$  is positive definite (PD) if for all non-zero vectors  $x \in \mathbb{R}^n$ ,  $x^T A x > 0$
- Similarly, positive semidefinite if  $x^T A x \geq 0$ , negative definite if  $x^T A x < 0$  and negative semidefinite if  $x^T A x \leq 0$
- One important property of positive definite and negative definite matrices is that they are always full rank, and hence, invertible.
- **Gram matrix:** Given any matrix  $A \in \mathbb{R}^{m \times n}$ , matrix  $G = A^T A$  is always positive semidefinite. Further if  $m \geq n$ , then  $G$  is positive definite.



We often encounter the quadratic form which is the vector increment of a quadratic function the quadratic form with respect to the matrix  $A$  of a vector  $X$  then the matrix  $M$  is  $n \times n$  and the vector  $X$  is  $n \times 1$  is given by the real number  $X^T A X$  based on the quadratic forms of matrices we can classify them as positive definite negative definite positive semi definite and negative semi definite.

A matrix  $A$  is said to be positive definite if its quadratic form is greater than zero for any vector  $X$  similarly we can define it to be negative definite, a matrix  $A$  is positive semi definite if the quadratic form is greater than equal to 0 for any vector  $X$  note that equality will 0 also make called it but the important property of positive and negative definite matrices is that they are always full rank.

An implication of this is that  $A^{-1}$  always exists, for a matrix  $A$  which is of dimension  $m \times m$  one can define a special matrix called a gram matrix, the gram matrix is given by  $A^T A$  the semi property of the gram matrix is that it is always positive semi definite, moreover if the number of rows exceeds the number of columns in other words if  $m \geq n$  this means this few lines that  $G$  is positive definite.

### IIT Madras Production

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

[www.nptel.ac.in](http://www.nptel.ac.in)

Copyrights Reserved