Introduction to Machine Learning in Geosciences GEO371T/GEO398D.1

Linear Algebra

Mrinal K. Sen

September 5, 2023

Linear Algebra

Essential for Understanding ML algorithms

- Scalars, vectors, matrices, tensors
- Multiplying matrices/vectors
- Inverse, Span, Linear Independence
- Linear Trabsformation, SVD, PCA

Scalars

- Notation x : scalar; $\mathbf{x}$ vector; $\mathbf{A}$: Matrix, Tensors
- A scalar is a single number
- It is represented by x.
- Example, slope of a line, a natural number, density etc.

Vector

- An array of numbers
- Arranged in order
- Each no. identified by an index
- Vectors are shown in lower-case bold

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ . \\ . \\ . \\ x_n \end{bmatrix}$$

$$\mathbf{x}^T = \begin{bmatrix} x_1 & x_2 & x_3 & . & . & . & x_n \end{bmatrix}^T$$

- If each element is in R then $\mathbf{x} \in R^n$
- We think of vectors as points in space - – Each element gives a coordinate along an axis

Matrix

- A 2-D array of numbers
- Each element identified by two indices
- Denoted by bold typeface **A**
- Elements indicated as $A_{i,j}$: i= row index, j=column index

$$\mathbf{A} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

- $\mathbf{A} \in R^{mXn}$

Tensor

- Sometimes need an array with more than two axes
- An array arranged on a regular grid with variable number of axes is referred to as a tensor
- Denote a tensor with bold typeface $\mathbf{A}$
- Elements indicated as $A_{i,j,k}$

Transpose of a Matrix, Matrix Addition

- Mirror image across the principal diagonal

$$\mathbf{A} = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix} \Rightarrow \mathbf{A}^T = \begin{bmatrix} A_{11} & A_{21} & A_{31} \\ A_{12} & A_{22} & A_{32} \\ A_{13} & A_{23} & A_{33} \end{bmatrix}$$

- Vectors are matrices with a single column

$$\mathbf{x} = [x_1, x_2, x_3, ..., x_n]^T$$

- $\mathbf{C} = \mathbf{A} + \mathbf{B} \Rightarrow C_{ij} = A_{ij} + B_{ij}$, $\mathbf{A} and \mathbf{B}$ are of the same shape!
- $\mathbf{D} = a\mathbf{B} + c \Rightarrow D_{ij} = aB_{ij} + c$
- $\mathbf{C} = \mathbf{A} + \mathbf{b} \Rightarrow C_{ij} = A_{ij} + b_j$ $\mathbf{D} = a\mathbf{B} + c \Rightarrow D_{ij} = aB_{ij} + c$

Matrix Multiplication

- For product **C=AB** to be defined, **A** has to have the same no. of columns as the no. of rows of **B**.
- If **A** is of shape *mxn* and **B** is of shape *nxp* then matrix product C is of shape *mxp*

$$\mathbf{C} = \mathbf{AB} \Rightarrow C_{ij} = \sum_k A_{ik} B_{kj}.$$

- Note that the standard product of two matrices is not just the product of two individual elements.
- Dot product of two vectors **x** and **y** of same dimensionality is the matrix product $\mathbf{x}^T\mathbf{y}$.
- Conversely, matrix product **C=AB** can be viewed as computing *Cij* the dot product of row i of **A** and column j of **B**.

Matrix Multiplication

- Distibutivity: $\mathbf{A(B+C)=AB+BC}$.
- Associativity: $\mathbf{A(BC)=(AB)C}$.
- Not commutative $\mathbf{AB} \neq \mathbf{BA}$.
- $\mathbf{x}^T\mathbf{y} = \mathbf{y}^T\mathbf{x}$.
- $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$

Linear Transformation

- $Ax=b$
  - where $A \in \mathbb{R}^{n \times n}$ and $b \in \mathbb{R}^n$
  - More explicitly

$$\begin{aligned} A_{11}x_1 + A_{12}x_2 + ... + A_{1n}x_n &= b_1 \\ A_{21}x_1 + A_{22}x_2 + ... + A_{2n}x_n &= b_2 \\ A_{n1}x_1 + A_{m2}x_2 + ... + A_{n,n}x_n &= b_n \end{aligned}$$

n equations in n unknowns

$$A = \begin{bmatrix} A_{1,1} & \cdots & A_{1,n} \\ \vdots & \vdots & \vdots \\ A_{n,1} & \cdots & A_{nn} \end{bmatrix} \quad x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \quad b = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$$

n x n $\quad$ n x 1 $\quad$ n x 1

Can view A as a *linear transformation* of vector x to vector b

- Sometimes we wish to solve for the unknowns $x = \{x_1,..,x_n\}$ when A and b provide constraints

12

Identity and Inverse Matrices

- Matrix inversion is a powerful tool to analytically solve $A\boldsymbol{x}=\boldsymbol{b}$
- Needs concept of Identity matrix
- Identity matrix does not change value of vector when we multiply the vector by identity matrix
  - Denote identity matrix that preserves n-dimensional vectors as I_n
  - Formally $I_n \in \mathbb{R}^{n\times n}$ and $\forall \mathbf{x} \in \mathbb{R}^n, I_n \mathbf{x} = \mathbf{x}$
  - Example of I_3 $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Matrix Inverse

- Inverse of square matrix A defined as $A^{-1}A = I_n$
- We can now solve $A\boldsymbol{x}=\boldsymbol{b}$ as follows:

$$A\boldsymbol{x} = \boldsymbol{b}$$
$$A^{-1}A\boldsymbol{x} = A^{-1}\boldsymbol{b}$$
$$I_n\boldsymbol{x} = A^{-1}\boldsymbol{b}$$
$$\boldsymbol{x} = A^{-1}\boldsymbol{b}$$

- This depends on being able to find A^{-1}
- If A^{-1} exists there are several methods for finding it

Linear Algebra

Solving Simultaneous equations

- $\mathrm{A}\boldsymbol{x} = \boldsymbol{b}$

  where A is $(M+1)$ x $(M+1)$

  $\boldsymbol{x}$ is $(M+1)$ x 1: set of weights to be determined

  $\boldsymbol{b}$ is N x 1
- Two closed-form solutions

  1. Matrix inversion $\boldsymbol{x}=\mathrm{A}^{-1}\boldsymbol{b}$
  2. Gaussian elimination

Deep Learning Srihari

How many solutions for $\boldsymbol{Ax}=\boldsymbol{b}$ exist?

- System of equations with
  - n variables and m equations is

$$
\begin{aligned}
A_{11}x_1 + A_{12}x_2 + \ldots + A_{1n}x_n &= b_1 \\
A_{21}x_1 + A_{22}x_2 + \ldots + A_{2n}x_n &= b_2 \\
A_{m1}x_1 + A_{m2}x_2 + \ldots + A_{mn}x_n &= b_m
\end{aligned}
$$

- Solution is $\boldsymbol{x}=A^{-1}\boldsymbol{b}$
- In order for A^{-1} to exist $\boldsymbol{Ax}=\boldsymbol{b}$ must have *exactly* one solution for every value of $\boldsymbol{b}$
  - It is also possible for the system of equations to have *no solutions* or an *infinite no. of solutions* for some values of $\boldsymbol{b}$
    - It is not possible to have more than one but fewer than infinitely many solutions
  - If $\boldsymbol{x}$ and $\boldsymbol{y}$ are solutions then $\boldsymbol{z}=\alpha\,\boldsymbol{x} + (1\text{-}\alpha)\,\boldsymbol{y}$ is a solution for any real α

19

Span of a set of vectors

- Span of a set of vectors: set of points obtained by a *linear combination* of those vectors
  - A linear combination of vectors $\{\boldsymbol{v}^{(1)},\ldots,\boldsymbol{v}^{(n)}\}$ with coefficients c_i is $\boxed{\sum_i c_i \boldsymbol{v}^{(i)}}$
  - System of equations is $A\boldsymbol{x}=\boldsymbol{b}$
    - A column of A, i.e., $A_{:i}$ specifies travel in direction i
    - How much we need to travel is given by x_i
    - This is a linear combination of vectors $\boxed{A\boldsymbol{x} = \sum_i x_i A_{:,i}}$
  - Thus determining whether $A\boldsymbol{x}=\boldsymbol{b}$ has a solution is equivalent to determining whether $\boldsymbol{b}$ is in the span of columns of A
    - This span is referred to as *column space* or *range* of A

Conditions for a solution to $Ax=b$

- Matrix must be square, i.e., $m=n$ and all columns must be *linearly independent*
  - Necessary condition is $n \geq m$
    - For a solution to exist when $b \in \mathbb{R}^m$ we require the column space be all of $\mathbb{R}^m$
  - Sufficient Condition
    - If columns are linear combinations of other columns, column space is less than $\mathbb{R}^m$
      - Columns are linearly dependent or matrix is *singular*
    - For column space to encompass $\mathbb{R}^m$ at least one set of m *linearly independent* columns
- For non-square and singular matrices
  - Methods other than matrix inversion are used

Norms

- Used for measuring the size of a vector
- Norms map vectors to non-negative values
- Norm of vector $\boldsymbol{x}$ is distance from origin to $\boldsymbol{x}$
  - It is any function f that satisfies:

$$f(\boldsymbol{x}) = 0 \Rightarrow \boldsymbol{x} = 0$$

$$f(\boldsymbol{x}+\boldsymbol{y}) \leq f(\boldsymbol{x}) + f(\boldsymbol{y}) \quad \text{Triangle Inequality}$$

$$\forall \alpha \in \mathbb{R} \quad f(\alpha \boldsymbol{x}) = |\alpha| f(\boldsymbol{x})$$

L^P Norm

- Definition $$\|x\|_p = \left(\sum_i |x_i|^p\right)^{\frac{1}{p}}$$
- L^2 Norm
  - Called Euclidean norm, written simply as $\|x\|$
  - Squared Euclidean norm is same as $x^T x$
- L^1 Norm
  - Useful when 0 and non-zero have to be distinguished (since L^2 increases slowly near origin, e.g., $0.1^2=0.01$)
- L^∞ Norm $$\|x\|_\infty = \max_i |x_i|$$
  - Called max norm

Size of a Matrix

- Frobenius norm

$$\boxed{\|A\|_F = \left(\sum_{i,j} A_{i,j}^2\right)^{\frac{1}{2}}}$$

- It is analogous to L^2 norm of a vector

Special kinds of Matrices

- Diagonal Matrix
  - Mostly zeros, with non-zero entries in diagonal
  - $diag\,(\boldsymbol{v})$ is a square diagonal matrix with diagonal elements given by entries of vector $\boldsymbol{v}$
  - Multiplying $diag(\boldsymbol{v})$ by vector $\boldsymbol{x}$ only needs to scale each element x_i by v_i

$$\boxed{diag(\boldsymbol{v})\boldsymbol{x} = \boldsymbol{v} \odot \boldsymbol{x}}$$

- Symmetric Matrix
  - Is equal to its transpose: $A=A^T$
  - E.g., a distance matrix is symmetric with $A_{ij}=A_{ji}$

Special Kinds of Vectors

- Unit Vector
  - A vector with unit norm $\boxed{\|x\|_2 = 1}$
- Orthogonal Vectors
  - A vector $\boldsymbol{x}$ and a vector $\boldsymbol{y}$ are orthogonal to each other if $\boldsymbol{x}^T\boldsymbol{y}=0$
  - Vectors are at 90 degrees to each other
- Orthogonal Matrix
  - A square matrix whose rows are mutually orthonormal

$$A^{-1}=A^T$$

Matrix decomposition

- Matrices can be decomposed into factors to learn universal properties about them not discernible from their representation
  - E.g., from decomposition of integer into prime factors $12=2x2x3$ we can discern that
    - 12 is not divisible by 5 or
    - any multiple of 12 is divisible by 3
    - But representations of 12 in binary or decimal are different
- Analogously, a matrix is decomposed into Eigenvalues and Eigenvectors to discern universal properties

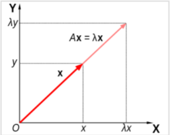
Deep Learning Srihari

Eigenvector

- An eigenvector of a square matrix A is a non-zero vector v such that multiplication by A only changes the scale of v

$$Av=\lambda v$$

  - The scalar λ is known as eigenvalue
- If v is an eigenvector of A, so is any rescaled vector sv. Moreover sv still has the same eigen value. Thus look for a unit eigenvector



Matrix A acts by stretching the vector x, not changing its direction, so x is an eigenvector of A.

Wikipedia

29

Eigenvalue and Characteristic Polynomial

- Consider $\boldsymbol{Av=w}$ $\quad A=\begin{bmatrix} A_{1,1} & \cdots & A_{1,n} \\ \vdots & \vdots & \vdots \\ A_{n,1} & \cdots & A_{nn} \end{bmatrix} \quad \boldsymbol{v}=\begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \quad \boldsymbol{w}=\begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}$
- If $\boldsymbol{v}$ and $\boldsymbol{w}$ are scalar multiples, i.e., if $A\boldsymbol{v}=\lambda\boldsymbol{v}$
  - then $\boldsymbol{v}$ is an eigenvector of the linear transformation A and the scale factor λ is the eigenvalue corresponding to the eigen vector
- This is the *eigenvalue equation* of matrix A
  - Stated equivalently as $(A-\lambda I)\boldsymbol{v}=0$
  - This has a non-zero solution if $|A-\lambda I|=0$ as
    - The polynomial of degree n can be factored as

$$|A-\lambda I| = (\lambda_1-\lambda)(\lambda_2-\lambda)\ldots(\lambda_n-\lambda)$$

    - The $\lambda_1, \lambda_2 \ldots \lambda_n$ are roots of the polynomial and are eigenvalues of A

Deep Learning Srihari

Example of Eigenvalue/Eigenvector

- Consider the matrix $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$
- Taking determinant of $(A\text{-}\lambda\mathrm{I})$, the char poly is

$$\boxed{|A-\lambda I| = \begin{bmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{bmatrix} = 3-4\lambda+\lambda^2}$$

- It has roots $\lambda=1$ and $\lambda=3$ which are the two eigenvalues of A
- The eigenvectors are found by solving for $\boldsymbol{v}$ in $A\boldsymbol{v}=\lambda\boldsymbol{v}$, which are

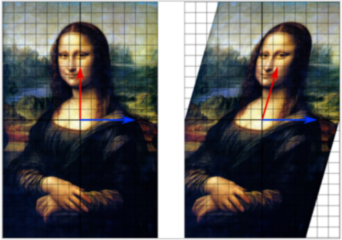
$$\boxed{v_{\lambda=1} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, v_{\lambda=3} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}}$$

31

Linear Algebra

Example of Eigen Vector

Vectors are grid points



In this shear mapping the red arrow changes direction but the blue arrow does not. The blue arrow is an eigenvector of this shear mapping because it doesn't change direction, and since its length is unchanged, its eigenvalue is 1.

Wikipedia

Eigendecomposition

- Suppose that matrix A has n linearly independent eigenvectors $\{\boldsymbol{v}^{(1)},\ldots,\boldsymbol{v}^{(n)}\}$ with eigenvalues $\{\lambda_1,\ldots,\lambda_n\}$
- Concatenate eigenvectors to form matrix V
- Concatenate eigenvalues to form vector $\lambda=[\lambda_1,\ldots,\lambda_n]$
- Eigendecomposition of A is given by

$$A = V\mathrm{diag}(\lambda)\, V^{-1}$$

Deep Learning Srihari

Decomposition of Symmetric Matrix

- Every real symmetric matrix A can be decomposed into real-valued eigenvectors and eigenvalues

$$A = Q\Lambda Q^T$$

where Q is an orthogonal matrix composed of eigenvectors of A: $\{v^{(1)},..,v^{(n)}\}$

orthogonal matrix: components are orthogonal or $v^{(i)T}v^{(j)}=0$

Λ is a diagonal matrix of eigenvalues $\{\lambda_1,..,\lambda_n\}$

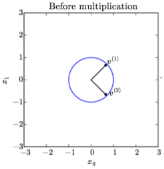
- We can think of A as scaling space by λ_i in direction $v^{(i)}$
  - See figure on next slide

34

Effect of Eigenvectors and Eigenvalues

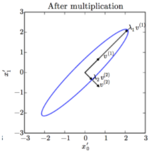
- Example of 2x2 matrix
- Matrix A with two orthonormal eigenvectors
  - $v^{(1)}$ with eigenvalue λ_1, $v^{(2)}$ with eigenvalue λ_2

Plot of unit vectors $\boldsymbol{u} \in \mathbb{R}^2$ (circle)



with two variables x_1 and x_2

Plot of vectors $A\boldsymbol{u}$ (ellipse)



Deep Learning Srihari

Linear Transformation

- $Ax=b$
  - where $A \in \mathbb{R}^{n \times n}$ and $b \in \mathbb{R}^n$
  - More explicitly

$$\begin{aligned} A_{11}x_1 + A_{12}x_2 + \ldots + A_{1n}x_n &= b_1 \\ A_{21}x_1 + A_{22}x_2 + \ldots + A_{2n}x_n &= b_2 \\ A_{n1}x_1 + A_{m2}x_2 + \ldots + A_{n,n}x_n &= b_n \end{aligned}$$

n equations in n unknowns

$$A = \begin{bmatrix} A_{1,1} & \cdots & A_{1,n} \\ \vdots & \vdots & \vdots \\ A_{n,1} & \cdots & A_{nn} \end{bmatrix} \quad x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \quad b = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$$

n x n $\qquad$ n x 1 $\qquad$ n x 1

Can view A as a *linear transformation* of vector x to vector b

- Sometimes we wish to solve for the unknowns $x = \{x_1,..,x_n\}$ when A and b provide constraints

12

What does eigendecomposition tell us?

- Many useful facts about the matrix obtained

1. Matrix is singular iff any of the eigenvalues are zero
2. Can be used to optimize quadratic expressions of the form

$$f(\boldsymbol{x}) = \boldsymbol{x}^T A \boldsymbol{x} \text{ subject to } ||\boldsymbol{x}||_2 = 1$$

Whenever $\boldsymbol{x}$ is equal to an eigenvector, f is equal to its eigenvalue

Max value of f is max eigen value, min value is min eigen value

Positive Definite Matrix

- A matrix whose eigenvalues are all positive is called *positive definite*
  - Positive or zero is called *positive semidefinite*
- If eigen values are all negative it is *negative definite*
- Positive definite matrices guarantee that $\boldsymbol{x}^T A \boldsymbol{x} \geq \mathbf{0}$

Singular Value Decomposition (SVD)

- Eigendecomposition has form:
  $A = V\text{diag}(\lambda)\, V^{-1}$
  - If A is not square eigendecomposition undefined
- SVD is a decomposition of the form $A = UDV^T$
- SVD is more general than eigendecomposition
  - Used with any matrix rather than symmetric ones
  - Every real matrix has a SVD (not so of eigen)

Trace of a Matrix

- Trace operator gives the sum of the elements along the diagonal

$$Tr(A) = \sum_{i,i} A_{i,i}$$

- Frobenius norm of a matrix can be represented as

$$\|A\|_F = \left(Tr(A)\right)^{\frac{1}{2}}$$

Determinant of a Matrix

- Determinant of a square matrix $\det(A)$ is a mapping to a scalar
- It is equal to the product of all eigenvalues of the matrix
- Measures how much multiplication by the matrix expands or contracts space