BLG 527E Machine Learning

FALL 2021-2022

Assoc. Prof. Yusuf Yaslan & Assist. Prof. Ayse Tosun

Linear Algebra for Machine Learning

Slides based on Mathematics for Machine Learning

Matrix

- **Definition 2.1** (Matrix)**.** With m, n ∈ N a real-valued (m, n) *matrix* A is an m.n-tuple of elements a_{ij} , i = 1,…,m, j = 1,…,n, which is ordered according to a rectangular scheme consisting of m rows and n columns:

- A = $\begin{matrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{matrix}$, $a_{ij} \in \mathbb{R}$

Matrix Addition

- A = $\begin{matrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{matrix}$, $a_{ij} \in \mathbb{R}$, m X n

- B = $\begin{matrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \vdots & \vdots & & \vdots \\ b_{m1} & b_{m2} & \dots & b_{mn} \end{matrix}$, $b_{ij} \in \mathbb{R}$, m X n

- A + B =

$$\begin{matrix} a_{11} + b_{11} & a_{12} + b_{12} & \dots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \dots & a_{2n} + b_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & \dots & a_{mn} + b_{mn} \end{matrix}$$

$\in \mathbb{R}^{mxn}$

Matrix Multiplication

- A $\in \mathbb{R}^{mxn}$, B $\in \mathbb{R}^{nxk}$
- C = AB $\in \mathbb{R}^{mxk}$ such that
- $c_{ij} = \sum_{l=1}^{n} a_{il} b_{lj}$
  $i = 1, .., m, j = 1, \dots, k$
- Not *cummutative*: AB ≠BA
- Associativity: A(BC) = (AB)C
- Distributivity: (A+B)C = AC + BC
  A(C+D) = AC + AD
- $I_m A = A I_n = A$

Identity matrix (I) is a matrix in $\mathbb{R}^{nxn}$ $(square\ matrix)$ $such\ that\ x_{ij} = 1\ for\ i = j$, and $x_{ij} = 0\ for\ i \neq j$

Multiplication by scalar also holds the associativity, distributivity properties.

Inverse

- Inverse of A is A^{-1}
- Not every matrix has an inverse. If inverse exists A is called *invertible/nonsingular,* otherwise *singular.*
- $A = \begin{matrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{matrix}$
- $A^{-1} = \dfrac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$ $iff\ determinant\ is\ not\ zero$

determinant of A

Transpose

- Transpose of A is A^T
- Write the rows of A as the columns in its transpose.
- Properties:
  - $AA^{-1} = I = A^{-1} A$
  - $(AB)^{-1} = B^{-1} A^{-1}$
  - $(A^T)^T = A$
  - $(A + B)^T = A^T + B^T$
  - $(AB)^T = B^T A^T$
- **Symmetric** matrix is a matrix A **if $\mathbf{A = A^T}$** Only *nxn* matrices (square) can be symmetric.

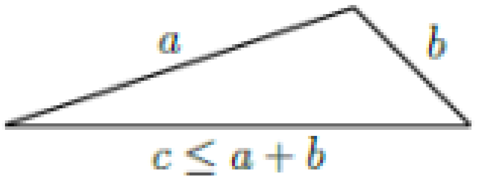
Vectors

- Norm of a vector is a function

$$\|.\| : V \rightarrow \mathbb{R}$$

which assigns each vector x its length $\|x\|$

The following properties:



- Absolutely homogeneous: $\|\lambda x\| = \|\lambda\|\|x\|$
- Triangle inequality: $\|x + y\| \leq \|x\| + \|y\|$
- Positive definite: $\|x\| \geq 0$ and $\|x\| = 0 \; then \; x = 0$

Orthogonal Matrix

- A square matrix is orthogonal iff its columns are orthonormal so that

$$AA^T = I = A^T A \text{ which implies that } A^{-1} = A^T$$

- Transformations with orthogonal matrices are special:

$$\|Ax\|^2 = \|x\|^2$$

Orthogonal matrices

- A = $\begin{bmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{bmatrix}$ (length of row is 1)

- Solve this by calculating $AA^T = I$

- A = $\begin{bmatrix} cos\theta & sin\theta \\ sin\theta & cos\theta \end{bmatrix}$ (also symmetric)

Matrix differentiation

- Let **y** be m-element vector (or mXn matrix) and **x** be an n-element vector.
- $\frac{\partial y}{\partial x} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \cdot & \cdot & & \cdot \\ \frac{\partial y_m}{\partial x_1} & \frac{\partial y_m}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix}$ (if x is a scalar, derivative of y wrt x is mX1 vector)
- Let where y is m×1, x is n×1, A is m×n
  - if **y=Ax,** then $\frac{\partial y}{\partial x}$= A (Solve it to see.)

Matrix/Vector differentiation

- Let the scalar α and y is m×1, x is n×1, A is m×n
  - if $\boldsymbol{\alpha} = \boldsymbol{y^T A x}$, then $\frac{\partial \boldsymbol{\alpha}}{\partial \boldsymbol{x}}$ = $\boldsymbol{y^T A}$ and $\frac{\partial \boldsymbol{\alpha}}{\partial \boldsymbol{y}}$ = $\boldsymbol{x^T A^T}$
- Let the scalar α and where x is n×1, A is n×n
  - α = $\boldsymbol{x^T A x}$, $\frac{\partial \alpha}{\partial \boldsymbol{x}}$ = $\boldsymbol{x^T}(\boldsymbol{A} + \boldsymbol{A^T})$
- For the special case where A is a *symmetric matrix*
  - α = $\boldsymbol{x^T A x}$, $\frac{\partial \alpha}{\partial \boldsymbol{x}}$ = $2\boldsymbol{x^T A}$