CS722/822: Machine Learning

Instructor: Jiangwen Sun
Computer Science Department

Basics of Linear Algebra

By $x \in \mathbb{R}^n$, we denote a vector with n entries

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

By $A \in \mathbb{R}^{m \times n}$ we denote a matrix with m rows and n columns

the jth column of A by a_j

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{bmatrix}$$

the ith row of A by a_i^T

Mathematical representation of data

- Examples
  - Training, validation, and test examples
- Features
  - Target(s) versus features
- Data matrix, data vector
  - Index of the examples

Matrix Multiplication

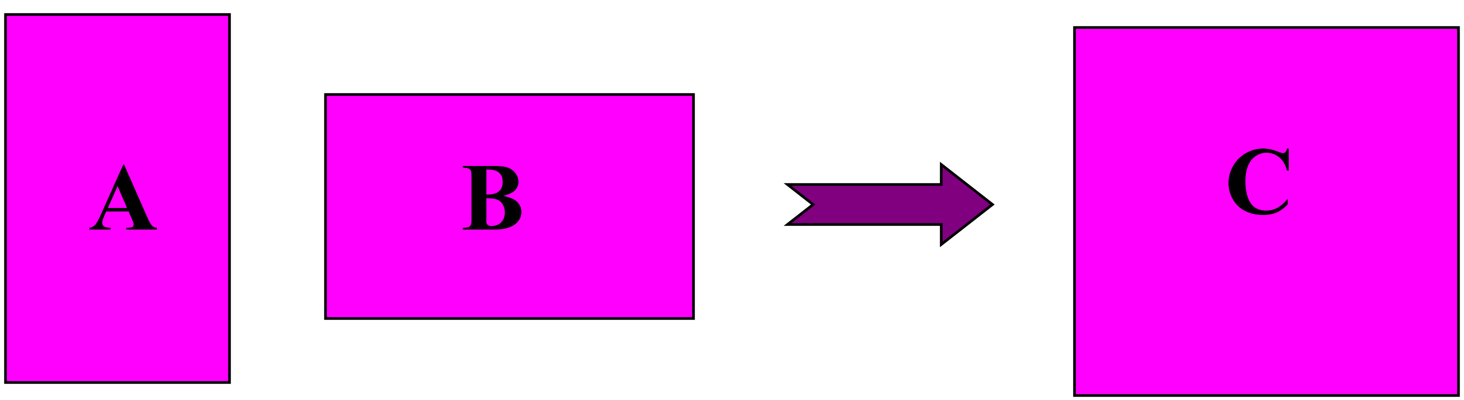
- The product of two matrices

$$A \in \mathbb{R}^{m\times n} \text{ and } B \in \mathbb{R}^{n\times p}$$

$$C = AB \in \mathbb{R}^{m\times p},$$

$$\text{where} \qquad C_{ij} = \sum_{k=1}^{n} A_{ik}B_{kj}\,.$$

- Special case: vector-vector product, matrix-vector product



Matrix Multiplication

C_{ij} is equal to the inner product of the ith row of A and the jth row of B

$$C = AB = \begin{bmatrix} — & a_1^T & — \\ — & a_2^T & — \\ & \vdots & \\ — & a_m^T & — \end{bmatrix} \begin{bmatrix} | & | & & | \\ b_1 & b_2 & \cdots & b_p \\ | & | & & | \end{bmatrix} = \begin{bmatrix} a_1^T b_1 & a_1^T b_2 & \cdots & a_1^T b_p \\ a_2^T b_1 & a_2^T b_2 & \cdots & a_2^T b_p \\ \vdots & \vdots & \ddots & \vdots \\ a_m^T b_1 & a_m^T b_2 & \cdots & a_m^T b_p \end{bmatrix}.$$

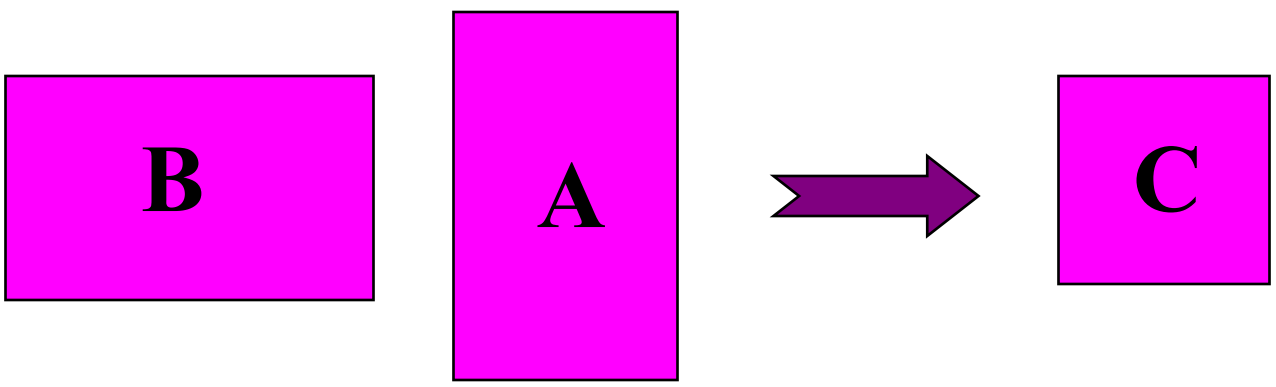
C is sum of outer products.

$$C = AB = \begin{bmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} — & b_1^T & — \\ — & b_2^T & — \\ & \vdots & \\ — & b_n^T & — \end{bmatrix} = \sum_{i=1}^{n} a_i b_i^T$$

Rules of Matrix Multiplication

- Matrix multiplication is associative: $(AB)C = A(BC)$.
- Matrix multiplication is distributive: $A(B + C) = AB + AC$.
- Matrix multiplication is, in general, *not* commutative;

$$AB \neq BA.$$

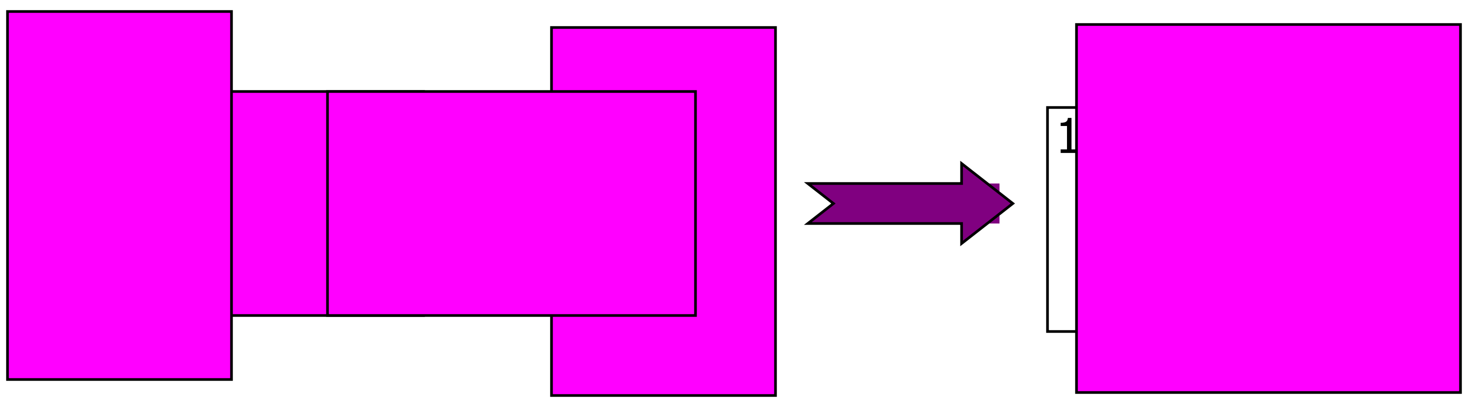


Orthogonal Matrix

$U \in \Re^{m\times m}$ is orthogonal, if and only if $U\,U^T = I_m$. (I_m is the identity matrix)

$\Rightarrow \mathrm{U}^{-1} = U^T$

The columns of $V \in \Re^{m\times n}\,(m > n)$ are orthormal, if and only if $V^T V = I$.



Square Matrix – EigenValue, EigenVector

a square matrix $A \in \mathbb{R}^{n \times n}$

(λ, x) is an eigen–pair of A, if and only if $Ax = \lambda x$.

λ is the eigenvalue

x is the eigenvector

write all the eigenvector equations

$AX = X\Lambda$ where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$

$$X \in \mathbb{R}^{n \times n} = \begin{bmatrix} | & | & & | \\ x_1 & x_2 & \cdots & x_n \\ | & | & & | \end{bmatrix}$$

$A = X\Lambda X^{-1}$ eigen-decomposition of A

Symmetric Matrix – EigenValue EigenVector

A is symmetric, if $A = A^T$

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

$A = X \Lambda X^T$ eigen-decomposition of A

.

$A \in \Re^{n \times n}$ is symmetric and positive semi-definite, if $x^T A x \geq 0$, for any $x \in \Re^n$.

$$\lambda_i \geq 0, \; i = 1, \cdots, n$$

$A \in \Re^{n \times n}$ is symmetric and positive definite, if $x^T A x > 0$, for any nonzero $x \in \Re^n$.

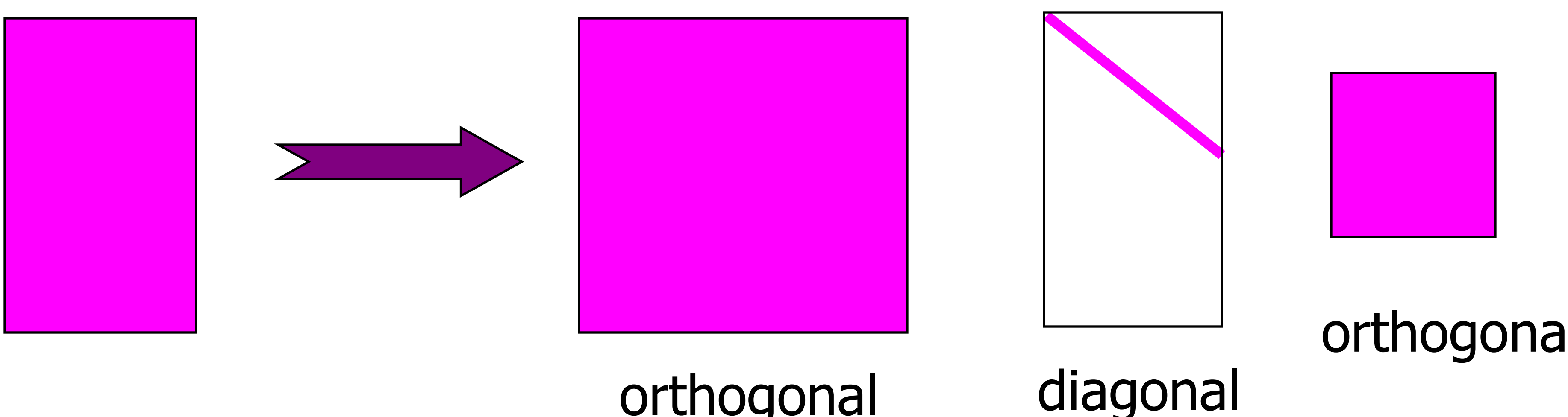
$$\lambda_i > 0, \; i = 1, \cdots, n$$

Singular Value Decomposition

Singular Value Decomposition (SVD): $A = U\Sigma V^T$, where $A \in \Re^{m\times n}$, $U \in \Re^{m\times m}$ and $V \in \Re^{n\times n}$ are orthogonal, and $\Sigma = diag(\sigma_1, \cdots, \sigma_r)$ is diagonal with $\sigma_1 \geq \cdots \geq \sigma_r$ and $r = \min(m,n)$.

$AA^T = U\Sigma\Sigma^T U^T$: U forms the eigenvectors of AA^T.

$A^T A = V\Sigma^T\Sigma V^T$: V forms the eigenvectors of $A^T A$.



Vector Norm

- l_2-norm

$$\|\mathbf{v}\|_2 = \sqrt{\sum_i v_i^2}$$

- l_1-norm

$$\|\mathbf{v}\|_1 = \sum_i |v_i|$$

- l_0-norm

$$\|\mathbf{v}\|_0 = \text{No. of none zero elements in } \mathbf{v}$$

Matrix Norms and Trace

Matrix norm:

2-norm: $\|A\|_2 =$ the square root of the largest eigenvalue of AA^T.

F-norm: $\|A\|_{\mathrm{F}} = \sqrt{\sum_{i,j} A_{ij}{}^2}$. Frobenius norm

1-norm: $\|A\|_1 = \sum_{i,j} |A_{ij}|$.

$\mathrm{trace}(A) = \sum_{i=1}^{m} A_{ii}$, for a square matrix A of size m by m.

$\|A\|_F^2 = \mathrm{trace}(AA^T) = \mathrm{trace}(A^T A)$, $\mathrm{trace}(AB) = \mathrm{trace}(BA)$.

$\|QA\|_F = \|A\|_F$, if Q has orthonormal columns.

References

- SC_linearAlg_basics.pdf (necessary)
- SVD_basics.pdf

loaded to Blackboard