신경망 네트워크와 수학적 기반

Matrix multiplication

- can multiply $m \times p$ matrix A and $p \times n$ matrix B to get $C = AB$:

$$C_{ij} = \sum_{k=1}^{p} A_{ik} B_{kj} = A_{i1} B_{1j} + \cdots + A_{ip} B_{pj}$$

for $i = 1, \ldots, m$, $j = 1, \ldots, n$

- to get C_{ij}: move along i-th row of A, j-th column of B
- example:

$$\begin{bmatrix} -1.5 & 3 & 2 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} -1 & -1 \\ 0 & -2 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 3.5 & -4.5 \\ -1 & 1 \end{bmatrix}$$

Special cases of matrix multiplication

- **scalar-vector product (with scalar on right) $x\alpha$**
- **inner product $a^T b$**
- **matrix-vector multiplication Ax**
- ***outer product* of m-vector a and n-vector b**

$$ab^T = \begin{bmatrix} a_1 b_1 & a_1 b_2 & \cdots & a_1 b_n \\ a_2 b_1 & a_2 b_2 & \cdots & a_2 b_n \\ \vdots & \vdots & & \vdots \\ a_m b_1 & a_m b_2 & \cdots & a_m b_n \end{bmatrix}$$

Matrix multiplication

Properties

- $(AB)C = A(BC)$, so both can be written ABC
- $A(B + C) = AB + AC$
- $(AB)^T = B^T A^T$
- $AI = A$ and $IA = A$
- *$AB = BA$ does not hold in general*

Matrix multiplication
Block matrices

- block matrices can be multiplied using the same formula, *e.g.*,

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} E & F \\ G & H \end{bmatrix} = \begin{bmatrix} AE + BG & AF + BH \\ CE + DG & CF + DH \end{bmatrix}$$

Matrix multiplication

Column interpretation

- denote columns of B by b_i:

$$B = \left[\begin{array}{cccc} b_1 & b_2 & \cdots & b_n \end{array} \right]$$

- then we have

$$\begin{aligned} AB &= A \left[\begin{array}{cccc} b_1 & b_2 & \cdots & b_n \end{array} \right] \\ &= \left[\begin{array}{cccc} Ab_1 & Ab_2 & \cdots & Ab_n \end{array} \right] \end{aligned}$$

- so AB is ‘batch’ multiply of A times columns of B

Multiple sets of linear equations

- given k systems of linear equations, with same $m \times n$ coefficient matrix

$$Ax_i = b_i, \quad i = 1, \ldots, k$$

- write in compact matrix form as $AX = B$
- $X = [x_1 \ \cdots \ x_k]$, $\mathrm{B} = [b_1 \ \cdots \ b_k]$

Matrix multiplication

Inner product interpretation

- with a_i^T the rows of A, b_j the columns of B, we have

$$AB = \begin{bmatrix} a_1^T b_1 & a_1^T b_2 & \cdots & a_1^T b_n \\ a_2^T b_1 & a_2^T b_2 & \cdots & a_2^T b_n \\ \vdots & \vdots & & \vdots \\ a_m^T b_1 & a_m^T b_2 & \cdots & a_m^T b_n \end{bmatrix}$$

- so matrix product is all inner products of rows of A and columns of B, arranged in a matrix

Matrix multiplication

Gram matrix

- let A be an $m \times n$ matrix with columns $a_1, \dots, a_n$
- the *Gram matrix* of A is

$$G = A^T A = \begin{bmatrix} a_1^T a_1 & a_1^T a_2 & \cdots & a_1^T a_n \\ a_2^T a_1 & a_2^T a_2 & \cdots & a_2^T a_n \\ \vdots & \vdots & \ddots & \vdots \\ a_n^T a_1 & a_n^T a_2 & \cdots & a_n^T a_n \end{bmatrix}$$

- Gram matrix gives all inner products of columns of A
- example: $G = A^T A = I$ means columns of A are orthonormal

Matrix multiplication

Composition of linear functions

- A is an $m \times p$ matrix, B is $p \times n$
- define $f : R^p \to R^m$ and $g : R^n \to R^p$ as

$$f(u) = Au, \quad g(v) = Bv$$

- f and g are linear functions
- *composition* of f and g is $h : \mathrm{R}^n \to \mathrm{R}^m$ with $h(x) = f(g(x))$
- we have

$$h(x) = f(g(x)) = A(Bx) = (AB)x$$

- composition of linear functions is linear
- associated matrix is product of matrices of the functions

Matrix multiplication

Second difference matrix

- D_n is $(n - 1) \times n$ difference matrix:

$$D_n x = (x_2 - x_1, \dots, x_n - x_{n-1})$$

- $D_{n\text{-}1}$ is $(n - 2) \times (n - 1)$ difference matrix:

$$D_n y = (y_2 - y_1, \dots, y_{n-1} - y_{n-2})$$

- $\Delta = D_{n-1} D_n$ is $(\mathrm{n} - 2) \times n$ second difference matrix:

$$\Delta \mathrm{x} = (x_1 - 2x_2 + x_3, x_2 - 2x_3 + x_4, \dots, x_{n-2} - 2x_{n-1} + x_n)$$

- for $n = 5$, $\Delta = D_{n-1}\mathrm{D}_n$ is

$$\begin{bmatrix} 1 & -2 & -1 & 0 & 0 \\ 0 & 1 & -2 & -1 & 0 \\ 0 & 0 & 1 & -2 & -1 \end{bmatrix} = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix}$$

Matrix powers

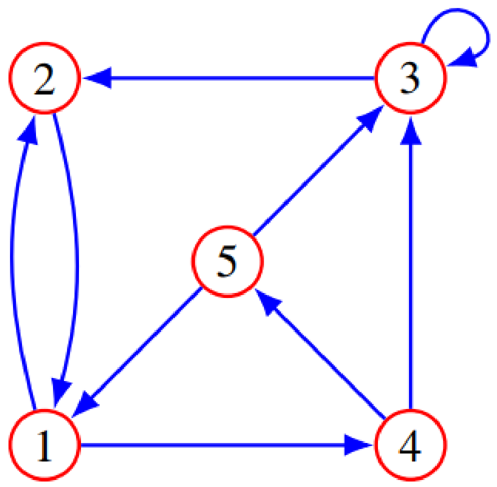
- for A square, A^2 means AA, and same for higher powers
- with convention $A^0 = I$ we have $A^k A^l = A^{k+l}$

Matrix multiplication
Directed graph

- $n \times n$ matrix A is adjacency matrix of directed graph:

$$A_{ij} = \begin{cases} 1 & \text{there is a edge from vertex } j \text{ to vertex } i \\ 0 & \text{otherwise} \end{cases}$$

- example:



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

Matrix multiplication

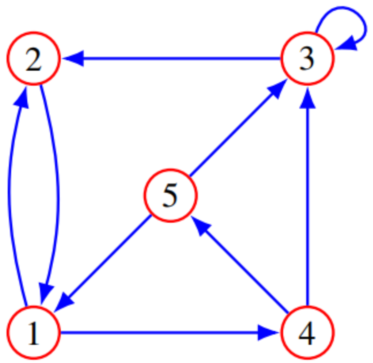
Paths in directed graph

- square of adjacency matrix:

$$(A^2)_{ij} = \sum_{k=1}^{n} A_{ik}A_{kj}$$

- $(A^2)_{ij}$ is number of paths of length 2 from j to i
- for the example,

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



e.g., there are two paths from 4 to 3 (via 3 and 5)

- more generally, $(A^\ell)ij$ = number of paths of length ℓ from j *to* i

Matrix multiplication

QR factorization

- $A = QR$ is called QR factorization of A
- factors satisfy $Q^TQ = I$, R upper triangular with positive diagonal entries