DEEP LEARNING

LINEAR ALGEBRA

LECTURE NOTES

LINEAR ALGEBRA - CHP 2

central in many machine learning alg.

MOTIVATING APPLICATION: SYSTEMS OF LINEAR EQUATIONS

Lin. eq. is a HYPERPLANE IN $\mathbb{R}^m$: set of points that satisfy a linear equation, e.g.

THE m-dimensional space

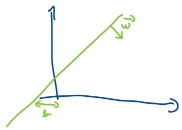
$$w_1 x_1 + w_2 x_2 + \dots + w_m x_m = b$$

- A point in an m-dimensional space is represented by a VECTOR $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}$
- We can represent the hyperplane with the vector of its weights $\vec{w} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_m \end{bmatrix}$ and the bias b
- $\vec{w}$ is itself a vector, orthogonal to the hyperplane.

OUTLINED LATER

$$\vec{w} \cdot \vec{x} = b$$

In $\mathbb{R}^2$, an hyperplane is just a line:



- Vector $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}$

- Matrix $A = \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix}$

indices may start from 0 or from 1!

$$(A^T)_{i,j} = A_{j,i}$$

- $\boxed{\text{Matrix Product:}}$ $C = A\,B$

$m \times n$ $n \times p$

$$C_{i,j} = \sum_k A_{i,k} \cdot B_{k,j}$$

MAKE EXAMPLE
MATRIX X VECTOR
MATRIX X MATRIX

EXAMPLES:

MATRIX · VECTOR

$$A \cdot \vec{b} = \underbrace{\begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \end{bmatrix}}_{\mathbb{R}^{2\times 3}} \cdot \underbrace{\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}}_{\mathbb{R}^{3\times 1}} = \underbrace{\begin{bmatrix} A_{11} b_1 + A_{12} b_2 + A_{13} b_3 \\ A_{21} b_1 + A_{22} b_2 + A_{23} b_3 \end{bmatrix}}_{\mathbb{R}^{2\times 1}}$$

MATRIX · MATRIX

$$A \cdot B = \underbrace{\begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \end{bmatrix}}_{\mathbb{R}^{2\times 3}} \cdot \underbrace{\begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \\ B_{31} & B_{32} \end{bmatrix}}_{\mathbb{R}^{3\times 2}} = \underbrace{\begin{bmatrix} A_{11}B_{11} + A_{12}B_{21} + A_{13}B_{31} & A_{11}B_{12} + A_{12}B_{22} + A_{13}B_{32} \\ A_{21}B_{11} + A_{22}B_{21} + A_{23}B_{31} & A_{21}B_{12} + A_{22}B_{22} + A_{23}B_{32} \end{bmatrix}}_{\mathbb{R}^{2\times 2}}$$

Properties of Matrix Product

Distributive: $A(B+C) = AB + AC$

Associative: $A(BC) = (AB)C$

DOT PROD.

~~Commutative~~ $AB \neq BA$ (but $x^T y = y^T x$)

$$(AB)^T = B^T A^T$$

- Element-Wise product (Hadamard) $A \odot B$

- Dot product (vectors) $\langle x, y \rangle = x^T y$

Systems of Linear Equations

$A\vec{x} = \vec{b}$ SAME AS

A: $m \times n$

$A_{1,:} x = b_1 \xrightarrow{\text{Same}} A_{11} x_1 + A_{12} x_2 + \dots + A_{1n} x_n = b_1$

$A_{2,:} x = b_2$

...

$A_{m,:} x = b_m$

- Identity Matrix: $\forall x \in \mathbb{R}^n, I_n x = x$ $\quad I_n = \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix}$ ($n \times n$)

Solve system of linear equations

$$Ax = b$$
$$A^{-1} A x = A^{-1} b$$
$$I_n x = A^{-1} b$$
$$x = A^{-1} b$$

- Matrix Inverse $A^{-1} A = I_n$

A^{-1} Don't always exist!

- A must be square, BUT NOT ALL SQUARE MATRICES ARE INVERTIBLE!
- Its columns must be independent (more details in chapter 2)

NORMS: how to measure the size of a vector.

L^p NORM: $\|\vec{x}\|_p = \sqrt[p]{\sum_i |x_i|^p}$

L^2 norm is the "EUCLIDEAN NORM" $\|\vec{x}\|_2$ or simply $\|\vec{x}\|$

L^1 norm: $\|\vec{x}\|_1 = \sum_i |x_i|$

SIZE of a Matrix: FROBENIUS norm $\|A\|_F = \sqrt{\sum_{i,j} A_{i,j}^2}$

The dot product between two vectors can be written as $\vec{x}^T \vec{y} = \|\vec{x}\| \|\vec{y}\| \cos\theta$ ← ANGLE BETWEEN $\vec{x}$ and $\vec{y}$

SPECIAL MATRICES AND VECTORS

DIAGONAL MATRICES: entries only in the diagonal $\quad diag(\vec{v}) = \begin{bmatrix} v_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & v_n \end{bmatrix}$

UNIT VECTOR: $\|\vec{x}\|_2 = 1$

ORTHOGONAL VECTORS $\vec{x}^T \vec{y} = 0$ $\quad$ ORTHONORMAL: ORTHOGONAL UNIT VECTORS

ORTHOGONAL MATRIX: rows and columns are mutually orthonormal $\quad A^T A = A A^T = I \Rightarrow A^{-1} = A^T$

EIGENDECOMPOSITION: UNDERSTANDING MATRICES DECOMPOSING THEM IN THEIR CONSTITUENT PARTS

EIGENVECTOR OF A SQUARE MATRIX A: NON-ZERO vector $\vec{v}$ such that multiplication by A only changes its scale (and not its direction)

$$A\vec{v} = \underset{\text{EIGENVALUE}}{\lambda} \underset{\text{EIGENVECTOR}}{\vec{v}}$$

If $\vec{v}$ is an eigenvector of A, then all $s\vec{v}$ for $s \in \mathbb{R}, s \neq 0$ are. We usually seek UNIT eigenvectors.

Suppose A has n eigenvectors. EIGENDECOMPOSITION: $A = \underset{\text{ONE EIGENVECTOR PER COLUMN}}{V} \, diag(\underset{\text{vector of eigenvalues}}{\vec{\lambda}}) V^{-1}$. Not all matrices have an eigendecomposition

SINGULAR VALUE DECOMPOSITION (SVD

More general than eigendecomposition. EVERY REAL MATRIX HAS AN SVD.

$$\underbrace{A}_{\mathbb{R}^{m\times n}} = \underbrace{U}_{\substack{\mathbb{R}^{m\times m} \\ \text{ORTHOGONAL}}} \ \underbrace{D}_{\mathbb{R}^{m\times n},\ \text{diagonal}} \ \underbrace{V^T}_{\mathbb{R}^{n\times n},\ \text{ORTHOGONAL}}$$

MOORE-PENROSE PSEUDOINVERSE

Generalization of inverse to rectangular matrices.

TRACE: the sum of the elements on the diagonal

EXERCISE

Assume $A = V \operatorname{diag}(\vec{\lambda})\, V^{-1}$

Does there exist an eigendecomposition for A^2? If so, what does it look like?

$A^2 = ?$