Linear Algebra

Hee-il Hahn

Professor
Department of Information and Communications Engineering
Hankuk University of Foreign Studies
hihahn@hufs.ac.kr

Norms

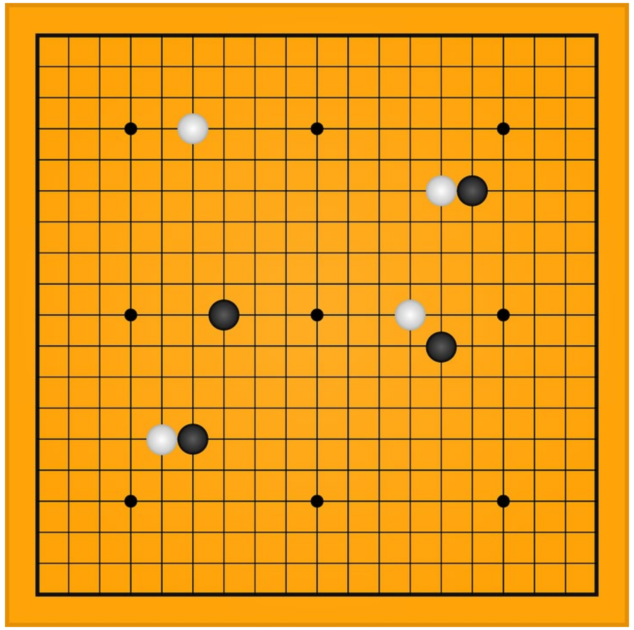
- L^p Norms

$$\|\boldsymbol{x}\|_p = \left(\sum_i |x_i|^p\right)^{\frac{1}{p}} \quad for\ \ p \in \boldsymbol{Z},\ p \geq 1$$

$$where\ \ \boldsymbol{x} = [x_1 \quad \cdots \quad x_n]^T$$

  - $L^1\ norm$: Metropolitan distance $\rightarrow \|\boldsymbol{x}\|_1 = |x_1| + |x_2| + |x_3|$
  - $L^2\ norm$: Euclidean distance $\rightarrow \|\boldsymbol{x}\|_2 = \sqrt{x_1^2 + x_2^2 + x_3^2}$
  - max $norm$: $\rightarrow \|\boldsymbol{x}\|_\infty = \max_i |x_i|$
  - Frobenius norm : measure the size of a matrix

$$\|\boldsymbol{A}\|_F = \sqrt{\sum_{i,j} A_{i,j}^2}$$



Norms – cont.

- Functions mapping vectors to non-negative values
- Measures the distance from the origin to the point x.
- More rigoriously, any function f that satisfies the following properties:
  - $f(\boldsymbol{x}) = 0 \;\; \rightarrow \;\; \boldsymbol{x} = 0$
  - $f(\boldsymbol{x} + \boldsymbol{y}) \leq f(\boldsymbol{x}) + f(\boldsymbol{y})$: triangle inequality
  - $for\ all\ \ \alpha \in R, \quad f(\alpha\boldsymbol{x}) = |\alpha| f(\boldsymbol{x})$

Special Kinds of Matrices and Vectors

- Diagonal matrix

$$\boldsymbol{D} = \begin{bmatrix} v_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & v_n \end{bmatrix} \qquad \boldsymbol{D}^{-1} = \begin{bmatrix} \frac{1}{v_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{v_n} \end{bmatrix}, \quad \boldsymbol{D}^{-1} \text{ exists only if } v_i \neq 0$$

- Symmetric matrix $\boldsymbol{A} = \boldsymbol{A}^T$
- Skew-symmetric matrix $\boldsymbol{A} = -\boldsymbol{A}^T$
- Orthogonal matrix $\boldsymbol{A}^T\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}^T = \boldsymbol{I} \ \rightarrow \ \boldsymbol{A}^{-1} = \boldsymbol{A}^T$
- Hermitian matrix $\overline{\boldsymbol{A}}^T = \boldsymbol{A} \ \ or \ \ \boldsymbol{A}^T = \overline{\boldsymbol{A}}$
- Skew-hermitian matrix $\overline{\boldsymbol{A}}^T = -\boldsymbol{A} \ \ or \ \ \boldsymbol{A}^T = -\overline{\boldsymbol{A}}$
- Unitary matrix $\overline{\boldsymbol{A}}^T = \boldsymbol{A}^{-1}$
- Unit vector : a vector with unit norm $\|\boldsymbol{x}\|_2 = 1$

Special Kinds of Matrices and Vectors – cont.

- Diagonal matrix

$$\boldsymbol{D} = \begin{bmatrix} 1.5 & 0 & 0 \\ 0 & 2.1 & 0 \\ 0 & 0 & 3.3 \end{bmatrix} \qquad \boldsymbol{D}^{-1} = \begin{bmatrix} \frac{1}{1.5} & 0 & 0 \\ 0 & \frac{1}{2.1} & 0 \\ 0 & 0 & \frac{1}{3.3} \end{bmatrix}, \quad \boldsymbol{D}^{-1} \text{ exists only if } v_i \neq 0$$

- Symmetric matrix $\boldsymbol{A} = \boldsymbol{A}^T$
- Skew-symmetric matrix $\boldsymbol{A} = -\boldsymbol{A}^T$
- Orthogonal matrix $\boldsymbol{A}^T\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}^T = \boldsymbol{I} \;\rightarrow\; \boldsymbol{A}^{-1} = \boldsymbol{A}^T$

$$\begin{bmatrix} -3 & 1 & 5 \\ 1 & 0 & -2 \\ 5 & -2 & 4 \end{bmatrix}, \quad \begin{bmatrix} 0 & 9 & -12 \\ -9 & 0 & 20 \\ 12 & -20 & 0 \end{bmatrix}, \quad \begin{bmatrix} \frac{2}{3} & \frac{1}{3} & \frac{2}{3} \\ -\frac{2}{3} & \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} & -\frac{2}{3} \end{bmatrix}$$

Special Kinds of Matrices and Vectors – cont.

- Hermitian matrix $\bar{\boldsymbol{A}}^T = \boldsymbol{A}$ or $\boldsymbol{A}^T = \bar{\boldsymbol{A}}$
- Skew-hermitian matrix $\bar{\boldsymbol{A}}^T = -\boldsymbol{A}$ or $\boldsymbol{A}^T = -\bar{\boldsymbol{A}}$
- Unitary matrix $\bar{\boldsymbol{A}}^T = \boldsymbol{A}^{-1}$
- Unit vector : a vector with unit norm $\|\boldsymbol{x}\|_2 = 1$

$$\mathbf{A} = \begin{bmatrix} 4 & 1-3i \\ 1+3i & 7 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 3i & 2+i \\ -2+i & -i \end{bmatrix} \quad \mathbf{C} = \begin{bmatrix} \frac{1}{2}i & \frac{1}{2}\sqrt{3} \\ \frac{1}{2}\sqrt{3} & \frac{1}{2}i \end{bmatrix}$$

Unit vector : $v = \left[\frac{1}{2}, \frac{1}{\sqrt{2}}, \frac{1}{2}\right]$

Eigendecomposition

- Decompose a matrix into a set of eigenvectors and eigenvalues.

$$\boldsymbol{A}\boldsymbol{v} = \lambda \boldsymbol{v} \qquad \lambda\text{: eigen value } \ \boldsymbol{v}\text{: eigen vector}$$

$$\boldsymbol{v}^T \boldsymbol{A} = \lambda \boldsymbol{v}^T \qquad \text{left eigen vector}$$

- Suppose $\boldsymbol{A}$ has linearly independent eigenvectors $\{v^{(1)}, \cdots, v^{(n)}\}$ with corresponding eigenvalues $\{\lambda_1, \cdots, \lambda_n\}$

Eigen decomposition : $\boldsymbol{A} = \boldsymbol{V}\, diag(\boldsymbol{\lambda})\, \boldsymbol{V}^{-1}$, 여기서 $\boldsymbol{V} = [v^{(1)}, \cdots, v^{(n)}]$

$(\because \ \boldsymbol{AV} = \boldsymbol{V}\, diag(\boldsymbol{\lambda}))$

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

$$\boldsymbol{A} = \boldsymbol{Q}\, diag(\boldsymbol{\lambda}) \boldsymbol{Q}^T$$

Eigendecomposition – cont.

- $A = \begin{bmatrix} 6 & -3 \\ 4 & -1 \end{bmatrix}$ 일때 eigen value, eigenvector 계산

- $Step\ 1$:

  $$det(\lambda I - A) = \begin{vmatrix} \lambda - 6 & 3 \\ -4 & \lambda + 1 \end{vmatrix} = \lambda^2 - 5\lambda + 6 = 0 \rightarrow \lambda = 3, 2$$

- $Step\ 2$:
  - $i)\ \lambda_1 = 3$:

    $$\lambda_1 I - A = \begin{bmatrix} -3 & 3 \\ -4 & 4 \end{bmatrix} \rightarrow \begin{bmatrix} -3 & 3 \\ -4 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 0 \rightarrow x - y = 0$$

    $$\therefore v^{(1)} = [1, \quad 1]^T$$

  - $ii)\ \lambda_2 = 2$: $v^{(2)} = [3, \quad 4]^T$

Eigendecomposition – cont.

- $Step\ 2$:
  - $ii)\ \lambda_2 = 2$:

$$\lambda_2 I - A = \begin{bmatrix} -4 & 3 \\ -4 & 3 \end{bmatrix} \rightarrow \begin{bmatrix} -4 & 3 \\ -4 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 0 \rightarrow -4x + 3y = 0$$

$$\therefore v^{(2)} = [3, \quad 4]^T$$

Eigendecomposition – cont.

- $\{v^{(1)}, v^{(2)}\}$: linearly independent eigenvectors with corresponding eigenvalues $\{3, 2\}$

  - Eigen decomposition : $\boldsymbol{A} = \boldsymbol{V} diag(\boldsymbol{\lambda})\boldsymbol{V}^{-1}$

$\begin{bmatrix} 6 & -3 \\ 4 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 1 & 4 \end{bmatrix}^{-1}$ 그런데, $\begin{bmatrix} 1 & 3 \\ 1 & 4 \end{bmatrix}^{-1} = \begin{bmatrix} 4 & -3 \\ -1 & 1 \end{bmatrix}$이므로

$\begin{bmatrix} 6 & -3 \\ 4 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 4 & -3 \\ -1 & 1 \end{bmatrix}$로 나타낼 수 있다.

  - Diagonalization: $diag(\boldsymbol{\lambda}) = \boldsymbol{V}^{-1}\ \boldsymbol{AV}$

$\begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 4 & -3 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 6 & -3 \\ 4 & -1 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 1 & 4 \end{bmatrix}$

Eigendecomposition – cont.

- $A = \begin{bmatrix} 5 & 2 \\ 2 & 2 \end{bmatrix}$ 일때 eigen values, eigenvectors 계산

- $Step\ 1$:

  $det(\lambda I - A) = \begin{vmatrix} \lambda - 5 & -2 \\ -2 & \lambda - 2 \end{vmatrix} = \lambda^2 - 7\lambda + 6 = 0 \rightarrow \lambda = 1, 6$

- $Step\ 2$:
  - $i)\ \lambda_1 = 1$:

    $\lambda_1 I - A = \begin{bmatrix} -4 & -2 \\ -2 & -1 \end{bmatrix} \rightarrow \begin{bmatrix} -4 & -2 \\ -2 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 0 \rightarrow 2x + y = 0$

    $\therefore v^{(1)} = [1, \quad -2]^T$

  - $ii)\ \lambda_2 = 6$:

Eigendecomposition – cont.

- Step 2 :
  - $ii)$ $\lambda_2 = 6$:

$$\lambda_2 I - A = \begin{bmatrix} 1 & -2 \\ -2 & 4 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & -2 \\ -2 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 0 \rightarrow x - 2y = 0$$

$$\therefore v^{(2)} = [2, \quad 1]^T$$

- Step 3 : $\boldsymbol{A} = \boldsymbol{V}\, diag(\boldsymbol{\lambda})\, \boldsymbol{V}^{-1} = \begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 6 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix}^{-1}$

$$\left(\begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix}^{-1} = \frac{1}{5} \begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix} \right) \qquad = \frac{1}{5} \begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 6 \end{bmatrix} \begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 6 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix} = \boldsymbol{Q}\, diag(\boldsymbol{\lambda})\, \boldsymbol{Q}^T$$

Eigendecomposition – cont.

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

$$\bar{x}^T A x = \lambda \bar{x}^T x \ \rightarrow \ \lambda = \frac{\bar{x}^T A x}{\bar{x}^T x} \quad \left(\because \ \bar{x}^T A x = x^T A^T \bar{x} = \overline{\bar{x}^T A x}, \ (AB)^T = B^T A^T\right)$$

  - A: symmetric $\Rightarrow \langle Av, w\rangle = \langle v, Aw\rangle$

    $\Rightarrow \langle Av, w\rangle = (Av)^T w = v^T A^T w = \langle v, A^T w\rangle = \langle v, Aw\rangle$

  - A: symmetric and v, w: eigenvectors with different eigenvalue

    $\Rightarrow \langle v, w\rangle = 0$

    (proof): suppose $Av = \lambda_1 v$ and $Aw = \lambda_2 w$

    $\Rightarrow \lambda_1\langle v, w\rangle = \langle \lambda_1 v, w\rangle = \langle Av, w\rangle = \langle v, Aw\rangle = \langle v, \lambda_2 w\rangle = \lambda_2\langle v, w\rangle$

  $\Rightarrow$ eigenvectors are orthogonal $\Rightarrow \boldsymbol{Q}^T = \boldsymbol{Q}^{-1}$

Quadratic Forms

$$
\begin{aligned}
Q = \mathbf{x}^{\mathsf{T}}\mathbf{A}\mathbf{x} &= \sum_{j=1}^{n}\sum_{k=1}^{n} a_{jk}x_jx_k \\
&= \quad a_{11}x_1^2 \quad + a_{12}x_1x_2 + \cdots + a_{1n}x_1x_n \\
&\quad + a_{21}x_2x_1 + a_{22}x_2^2 \quad + \cdots + a_{2n}x_2x_n \\
&\quad + \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
&\quad + a_{n1}x_nx_1 + a_{n2}x_nx_2 + \cdots + a_{nn}x_n^2.
\end{aligned}
$$

Quadratic Form. Symmetric Coefficient Matrix

Let

$$\mathbf{x}^{\mathsf{T}}\mathbf{A}\mathbf{x} = [x_1 \quad x_2]\begin{bmatrix} 3 & 4 \\ 6 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 3x_1^2 + 4x_1x_2 + 6x_2x_1 + 2x_2^2 = 3x_1^2 + 10x_1x_2 + 2x_2^2.$$

Here $4 + 6 = 10 = 5 + 5$. From the corresponding ***symmetric*** matrix $\mathbf{C} = [c_{jk}]$, where $c_{jk} = \frac{1}{2}(a_{jk} + a_{kj})$, thus $c_{11} = 3$, $c_{12} = c_{21} = 5$, $c_{22} = 2$, we get the same result; indeed,

$$\mathbf{x}^{\mathsf{T}}\mathbf{C}\mathbf{x} = [x_1 \quad x_2]\begin{bmatrix} 3 & 5 \\ 5 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 3x_1^2 + 5x_1x_2 + 5x_2x_1 + 2x_2^2 = 3x_1^2 + 10x_1x_2 + 2x_2^2. \quad \blacksquare$$

Quadratic Forms – cont.

- $f(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x}$ $\quad (\boldsymbol{A} = \boldsymbol{Q}\Lambda\boldsymbol{Q}^T)$이용하면

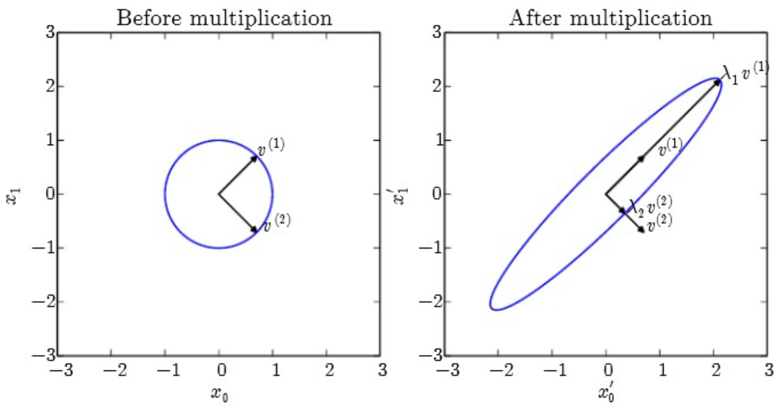
$$= \boldsymbol{x}^T\boldsymbol{Q}\Lambda\boldsymbol{Q}^T\boldsymbol{x} = \boldsymbol{x}^T\boldsymbol{Q}\Lambda\boldsymbol{Q}^T\boldsymbol{x} = (\boldsymbol{Q}^T\boldsymbol{x})^{\boldsymbol{T}}\Lambda(\boldsymbol{Q}^T\boldsymbol{x}) = \boldsymbol{y}^T\Lambda\boldsymbol{y}, \quad (\boldsymbol{y} = \boldsymbol{Q}^T\boldsymbol{x}) \text{ 이용}$$

$$= [y_1 \quad y_2]\begin{bmatrix}\lambda_1 & 0 \\ 0 & \lambda_2\end{bmatrix}\begin{bmatrix}y_1 \\ y_2\end{bmatrix}$$

$$= \lambda_1 y_1^2 + \lambda_2 y_2^2$$

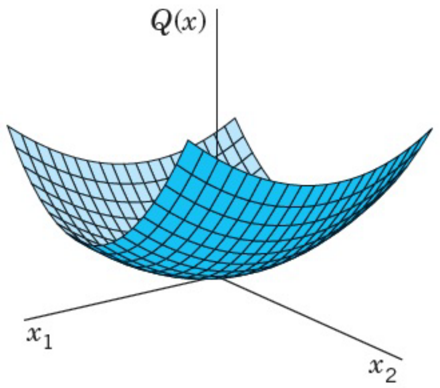
  - Optimization of quadratic form $f(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}$ subject to $\|\boldsymbol{x}\|_2 = 1$

$$\|\boldsymbol{x}\|_2 = 1 \equiv \sqrt{x_1^2 + x_2^2} = 1 \rightharpoonup \boldsymbol{x}^T\boldsymbol{x} = \mathbf{1}$$

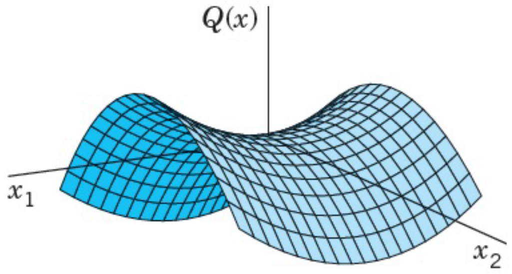


Quadratic Forms – cont.

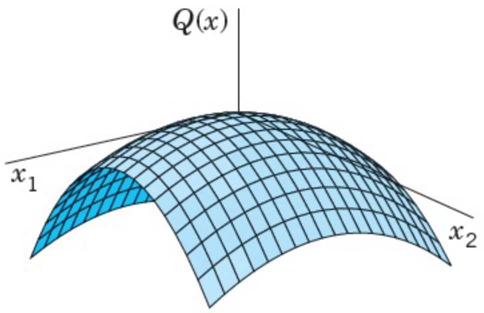
- Effect of eigenvectors and eivenvalues
  - Matrix is singular iff any of the eigenvalues are 0. (error)
  - Matrix whose eigenvalues are all positive ⇨ positive definite
  - *cf*: positive semidefinite, negative (semi)definite.
  - Positive definite matrix satisfying $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} = 0 \;\Rightarrow\; \boldsymbol{x} = 0$



(a) Positive definite form



(c) Indefinite form



(b) Negative definite form

Singular value decomposition

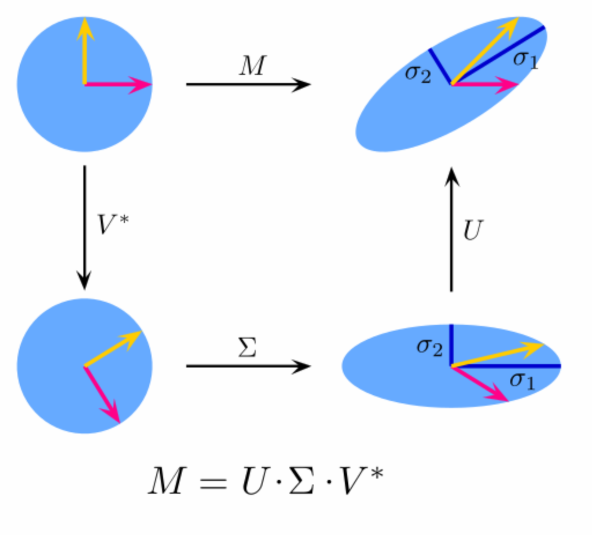
- *More general than eigenvalue decomposition*
- $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^T$ where $\boldsymbol{A}_{m\times n}$, $\boldsymbol{U}_{m\times m}$, $\boldsymbol{D}_{m\times n}$, $\boldsymbol{V}_{n\times n}$

  $\boldsymbol{U}, \boldsymbol{V}$: orthogonal matrix

columns of $\boldsymbol{U}$: left-singular vectors, eigenvectors of $\boldsymbol{A}\boldsymbol{A}^T$

columns of $\boldsymbol{V}$: right-singular vectors, eigenvectors of $\boldsymbol{A}^T\boldsymbol{A}$

nonzero singular values of $\boldsymbol{A}$: square roots of eigenvalues of $\boldsymbol{A}\boldsymbol{A}^T$ or $\boldsymbol{A}^T\boldsymbol{A}$



Singular value decomposition – cont'd

- *SVD Example*

$$\boldsymbol{A} = \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{bmatrix}$$

$$\boldsymbol{A}\boldsymbol{A}^T = \begin{bmatrix} 17 & 8 \\ 8 & 17 \end{bmatrix} \quad \boldsymbol{A}^T\boldsymbol{A} = \begin{bmatrix} 13 & 12 & 2 \\ 12 & 13 & -2 \\ 2 & -2 & 8 \end{bmatrix}$$

$$det(\boldsymbol{A}\boldsymbol{A}^T - \lambda\boldsymbol{I}) = \lambda^2 - 34\lambda + 225 = (\lambda - 25)(\lambda - 9) = 0$$

Singular values: $\sigma_1 = 5, \sigma_2 = 3$

- $\boldsymbol{A}^T\boldsymbol{A}$ is symmetric $\longrightarrow$ orthogonal eigenvectors
- For $\lambda = 25$,

$$\boldsymbol{A}^T\boldsymbol{A} - 25\boldsymbol{I} = \begin{bmatrix} -12 & 12 & 2 \\ 12 & -12 & -2 \\ 2 & -2 & -17 \end{bmatrix} \xrightarrow[row\ reduced]{} \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

Singular value decomposition – cont'd

- *SVD Example*
  - $\boldsymbol{A}^T\boldsymbol{A}$ is symmetric $\longrightarrow$ orthogonal eigenvectors
  - For $\lambda = 25$,

$$\boldsymbol{A}^T\boldsymbol{A} - 25\boldsymbol{I} = \begin{bmatrix} -12 & 12 & 2 \\ 12 & -12 & -2 \\ 2 & -2 & -17 \end{bmatrix} \xrightarrow[\text{row reduced}]{} \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

$$(\boldsymbol{A}^T\boldsymbol{A} - 25\boldsymbol{I})v = 0 \rightarrow \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0 \rightarrow \begin{matrix} x_1 - x_2 = 0 \\ x_3 = 0 \end{matrix}, \; x_1 = 1, x_2 = 1, x_3 = 0$$

$v^T = [1 \quad 1 \quad 0]$ 를 *normalize*시키면 $\quad v_1^T = \left[\frac{1}{\sqrt{2}} \quad \frac{1}{\sqrt{2}} \quad 0\right]$

Singular value decomposition – cont'd

- *SVD Example*
  - $\boldsymbol{A}^T\boldsymbol{A}$ is symmetric $\longrightarrow$ orthogonal eigenvectors
  - For $\lambda = 9$,

$$\boldsymbol{A}^T\boldsymbol{A} - 9\boldsymbol{I} = \begin{bmatrix} 4 & 12 & 2 \\ 12 & 4 & -2 \\ 2 & -2 & -1 \end{bmatrix} \xrightarrow[row\ reduced]{} \begin{bmatrix} 1 & 0 & -\frac{1}{4} \\ 0 & 1 & \frac{1}{4} \\ 0 & 0 & 0 \end{bmatrix}$$

$$(\boldsymbol{A}^T\boldsymbol{A} - 9\boldsymbol{I})v_2 = 0 \ \rightarrow \begin{bmatrix} 1 & 0 & -\frac{1}{4} \\ 0 & 1 & \frac{1}{4} \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0 \ \rightarrow \ \begin{aligned} x_1 - \tfrac{1}{4}x_3 &= 0 \\ x_2 + \tfrac{1}{4}x_3 &= 0 \end{aligned} \qquad x_1 = 1, x_2 = -1, x_3 = 4$$

$$v^T = [1 \quad -1 \quad 4] \text{를 } normalize\text{시키면} \quad v_2^T = \begin{bmatrix} \frac{1}{\sqrt{18}} & -\frac{1}{\sqrt{18}} & \frac{4}{\sqrt{18}} \end{bmatrix}$$

Singular value decomposition – cont'd

- *SVD Example*
  - for the last eigenvector $v_3^T = [a \quad b \quad c]$,

$$v_1^T v_3 = 0,\ v_2^T v_3 = 0 \longrightarrow a + b = 0, \frac{2}{\sqrt{18}} a + \frac{4}{\sqrt{18}} c = 0$$

$$\longrightarrow \quad v_3^T = \begin{bmatrix} a & -a & -\frac{1}{2} a \end{bmatrix} \quad \longrightarrow \quad v_3^T = \begin{bmatrix} \frac{2}{3} & -\frac{2}{3} & -\frac{1}{3} \end{bmatrix}$$

  - $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^T = \boldsymbol{U} \begin{bmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{18}} & -\frac{1}{\sqrt{18}} & \frac{4}{\sqrt{18}} \\ \frac{2}{3} & -\frac{2}{3} & -\frac{1}{3} \end{bmatrix}$

  - Finally, $\boldsymbol{A}v_i = \sigma_i u_i$ and $\boldsymbol{A}^T u_i = \sigma_i v_i \longrightarrow u_i = \frac{1}{\sigma_i} \boldsymbol{A} v_i \quad \longrightarrow \quad \boldsymbol{U} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$

Quiz

- 다음 행렬을 eigendecomposition 하시오.

  (문제를 풀고 사진을 찍어서 3월 20일 오후 12시까지 이클래스에 제출하세요.)

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

강의 정리

- 머신러닝을 이해하는데 필요한 최소한의 Linear algebra 리뷰
  - Symmetric, Hermitian, orthogonal, unitary matrix 등의 구조 및 특성 설명
  - Eigendecomposition 내용 및 계산하는 방법 설명
  - Quadratic Form 설명
  - Singular value decomposition 예제를 통한 계산법 설명

Moore-Penrose pseudoinverse

- Matrix inverse is not defined for matrices that are not square.
- given $\boldsymbol{Ax} = \boldsymbol{y}$, get a left-inverse of $\boldsymbol{A}$ s.t. $\boldsymbol{x} = \boldsymbol{By}$ and $\boldsymbol{A}_{m \times n}$
  - If $m > n$, possibly no solution
  - If $m < n$, multiple possible solutions
- Pseudoinverse is defined as

$$\boldsymbol{A}^{+} = \lim_{\alpha \to 0} (\boldsymbol{A}^{T}\boldsymbol{A} + \alpha \boldsymbol{I})^{-1} \boldsymbol{A}^{T}$$

- Practically,

$$\boldsymbol{A}^{+} = \boldsymbol{V}\boldsymbol{D}^{+}\boldsymbol{U}^{T}$$

  - If $m < n$, solution $\boldsymbol{x} = \boldsymbol{A}^{+}\boldsymbol{y}$ has minimal Euclidean norm $\|\boldsymbol{x}\|_2$
  - If $m > n$, possibly no solution, or $\boldsymbol{x}$ s.t. $\boldsymbol{Ax}$ is as close as possible to $\boldsymbol{y}$ in terms of Euclidean norm $\|\boldsymbol{Ax} - \boldsymbol{y}\|_2$
  - $\boldsymbol{D}^{+}$ is obtained by taking the reciprocal of its nonzero elements and taking the transpose of the resulting matrix.

Moore-Penrose pseudoinverse – cont'd

- When $\boldsymbol{A}$ has more columns than rows, it provides one of the many possible solutions.

- Solution $\boldsymbol{x} = \boldsymbol{A}^{+}\boldsymbol{y}$ has minimal Euclidean norm among all possible solutions.

Trace operator

- Definition

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

- Example

$$\|\boldsymbol{A}\|_F = \sqrt{Tr(\boldsymbol{A}\boldsymbol{A}^T)}$$

$$Tr(\boldsymbol{A}) = Tr(\boldsymbol{A}^T)$$

$$Tr(\boldsymbol{ABC}) = Tr(\boldsymbol{CAB}) = Tr(\boldsymbol{BCA})$$

$$Tr(\boldsymbol{AB}) = Tr(\boldsymbol{BA})$$

(if $\boldsymbol{A}_{m\times n}$, $\boldsymbol{B}_{n\times m}$ then $\boldsymbol{AB}_{m\times m}$, $\boldsymbol{BA}_{n\times n}$)

Principal component analysis

- Suppose a collection of m points $\{\boldsymbol{x}^{(1)} \quad \ldots \quad \boldsymbol{x}^{(m)}\}$ in $\boldsymbol{R}^n$
  - lossy compression assumed.
  - for each point $\boldsymbol{x}^{(k)}$, corresponding code vector $\boldsymbol{c}^{(k)} \in \boldsymbol{R}^l$, $l \leq n$
  - Find some encoding function $f(\boldsymbol{x}) = \boldsymbol{c}$ and decoding function $\boldsymbol{x} \approx g\big(f(\boldsymbol{x})\big)$
  - Let $g(\boldsymbol{c}) = \boldsymbol{D}\boldsymbol{c}$ where $\boldsymbol{D} \in \boldsymbol{R}^{n\times l}$

    constraints: columns of $\boldsymbol{D}$ are orthonormal to each other.
  - For optimal code ⇨ minimize the distance between $\boldsymbol{x}$ and $g(\boldsymbol{c})$
  - i.e.

$$
\begin{aligned}
\boldsymbol{c}^* &= \underset{\boldsymbol{c}}{\operatorname{argmin}} \|\boldsymbol{x} - g(\boldsymbol{c})\|_2 \\
&= \underset{\boldsymbol{c}}{\operatorname{argmin}} \big(\boldsymbol{x} - g(\boldsymbol{c})\big)^T \big(\boldsymbol{x} - g(\boldsymbol{c})\big) \\
&= \underset{\boldsymbol{c}}{\operatorname{argmin}} \big(\boldsymbol{x}^T\boldsymbol{x} - \boldsymbol{x}^T g(\boldsymbol{c}) - g(\boldsymbol{c})^T\boldsymbol{x} + g(\boldsymbol{c})^T g(\boldsymbol{c})\big) \\
&= \underset{\boldsymbol{c}}{\operatorname{argmin}} \big(-2\boldsymbol{x}^T g(\boldsymbol{c}) + g(\boldsymbol{c})^T g(\boldsymbol{c})\big)
\end{aligned}
$$

Principal component analysis

$$
\begin{aligned}
\boldsymbol{c}^* &= \underset{\boldsymbol{c}}{\operatorname{argmin}} \|\boldsymbol{x} - g(\boldsymbol{c})\|_2 \\
&= \underset{\boldsymbol{c}}{\operatorname{argmin}} \big(\boldsymbol{x} - g(\boldsymbol{c})\big)^T \big(\boldsymbol{x} - g(\boldsymbol{c})\big) \\
&= \underset{\boldsymbol{c}}{\operatorname{argmin}} \big(\boldsymbol{x}^T\boldsymbol{x} - \boldsymbol{x}^T g(\boldsymbol{c}) - g(\boldsymbol{c})^T\boldsymbol{x} + g(\boldsymbol{c})^T g(\boldsymbol{c})\big) \\
&= \underset{\boldsymbol{c}}{\operatorname{argmin}} \big(-2\boldsymbol{x}^T g(\boldsymbol{c}) + g(\boldsymbol{c})^T g(\boldsymbol{c})\big) \\
&= \underset{\boldsymbol{c}}{\operatorname{argmin}} (-2\boldsymbol{x}^T\boldsymbol{D}\boldsymbol{c} + \boldsymbol{c}^T\boldsymbol{D}^T\boldsymbol{D}\boldsymbol{c}) \\
&= \underset{\boldsymbol{c}}{\operatorname{argmin}} (-2\boldsymbol{x}^T\boldsymbol{D}\boldsymbol{c} + \boldsymbol{c}^T\boldsymbol{c}) \qquad (\boldsymbol{D}^T\boldsymbol{D} = \boldsymbol{I})
\end{aligned}
$$

$\therefore\ \boldsymbol{c} = \boldsymbol{D}^T\boldsymbol{x}$ ⇨ $f(\boldsymbol{x}) = \boldsymbol{D}^T\boldsymbol{x}$ and

$r(\boldsymbol{x}) = g(f(\boldsymbol{x})) = \boldsymbol{D}\boldsymbol{D}^T\boldsymbol{x}$

Principal component analysis

- How to choose the encoding matrix $\boldsymbol{D} \in \boldsymbol{R}^{n\times l}$

$$\boldsymbol{D}^* = \underset{\boldsymbol{D}}{\operatorname{argmin}} \sqrt{\sum_{i,j}\left(x_j^{(i)} - r(\boldsymbol{x}^{(i)})_j\right)^2} \quad \text{subject to } \boldsymbol{D}^T\boldsymbol{D} = \boldsymbol{I}$$

Consider the case of $l = 1$.

$$\boldsymbol{d}^* = \underset{\boldsymbol{d}}{\operatorname{argmin}} \sum_i \left\|\boldsymbol{x}^{(i)} - \boldsymbol{d}\boldsymbol{d}^T\boldsymbol{x}^{(i)}\right\|_2^2 \quad \text{subject to } \|\boldsymbol{d}\|_2 = 1$$

$$= \underset{\boldsymbol{d}}{\operatorname{argmin}} \sum_i \left\|\boldsymbol{x}^{(i)} - \boldsymbol{d}^T\boldsymbol{x}^{(i)}\boldsymbol{d}\right\|_2^2 \quad \text{subject to } \|\boldsymbol{d}\|_2 = 1$$

$$= \underset{\boldsymbol{d}}{\operatorname{argmin}} \sum_i \left\|\boldsymbol{x}^{(i)} - \boldsymbol{x}^{(i)^T}\boldsymbol{d}\boldsymbol{d}\right\|_2^2 \quad \text{subject to } \|\boldsymbol{d}\|_2 = 1$$

$$= \underset{\boldsymbol{d}}{\operatorname{argmin}} \|X - X\boldsymbol{d}\boldsymbol{d}^T\|_F^2 \quad \text{subject to } \boldsymbol{d}^T\boldsymbol{d} = 1$$

$$= \underset{\boldsymbol{d}}{\operatorname{argmin}}\, Tr\left((X - X\boldsymbol{d}\boldsymbol{d}^T)^T(X - X\boldsymbol{d}\boldsymbol{d}^T)\right) \quad \text{subject to } \|\boldsymbol{d}\|_2 = 1$$

$$= \underset{\boldsymbol{d}}{\operatorname{argmin}}\left(Tr(\boldsymbol{X}^T\boldsymbol{X} - \boldsymbol{X}^T\boldsymbol{X}\boldsymbol{d}\boldsymbol{d}^T - \boldsymbol{d}\boldsymbol{d}^T\boldsymbol{X}^T\boldsymbol{X} + \boldsymbol{d}\boldsymbol{d}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{d}\boldsymbol{d}^T)\right)$$

Principal component analysis

- How to choose the encoding matrix $\boldsymbol{D} \in \boldsymbol{R}^{n \times l}$

$$\boldsymbol{d}^* = \underset{\boldsymbol{d}}{\operatorname{argmin}} \sum_i \left\| \boldsymbol{x}^{(i)} - \boldsymbol{d}\boldsymbol{d}^T \boldsymbol{x}^{(i)} \right\|_2^2 \quad \text{subject to } \|\boldsymbol{d}\|_2 = 1 \quad (= \ \boldsymbol{d}^T\boldsymbol{d} = 1 \ )$$

$$= \underset{\boldsymbol{d}}{\operatorname{argmin}} \left(-2Tr(\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{d}\boldsymbol{d}^T) + Tr(\boldsymbol{d}\boldsymbol{d}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{d}\boldsymbol{d}^T)\right)$$

$$= \underset{\boldsymbol{d}}{\operatorname{argmin}} \left(-2Tr(\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{d}\boldsymbol{d}^T) + Tr(\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{d}\boldsymbol{d}^T\boldsymbol{d}\boldsymbol{d}^T)\right) \quad \text{subject to } \boldsymbol{d}^T\boldsymbol{d} = 1$$

$$= \underset{\boldsymbol{d}}{\operatorname{argmin}} \left(-2Tr(\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{d}\boldsymbol{d}^T) + Tr(\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{d}\boldsymbol{d}^T)\right) \quad \text{subject to } \boldsymbol{d}^T\boldsymbol{d} = 1$$

$$= \underset{\boldsymbol{d}}{\operatorname{argmax}} \left(Tr(\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{d}\boldsymbol{d}^T)\right) \quad \text{subject to } \boldsymbol{d}^T\boldsymbol{d} = 1$$

$$= \underset{\boldsymbol{d}}{\operatorname{argmax}} \left(Tr(\boldsymbol{d}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{d})\right) \quad \text{subject to } \boldsymbol{d}^T\boldsymbol{d} = 1$$

⇨ optimal $\boldsymbol{d} = \boldsymbol{d}^* =$ eigenvector of $\boldsymbol{X}^T\boldsymbol{X}$ corresponding to the largest eigenvalue