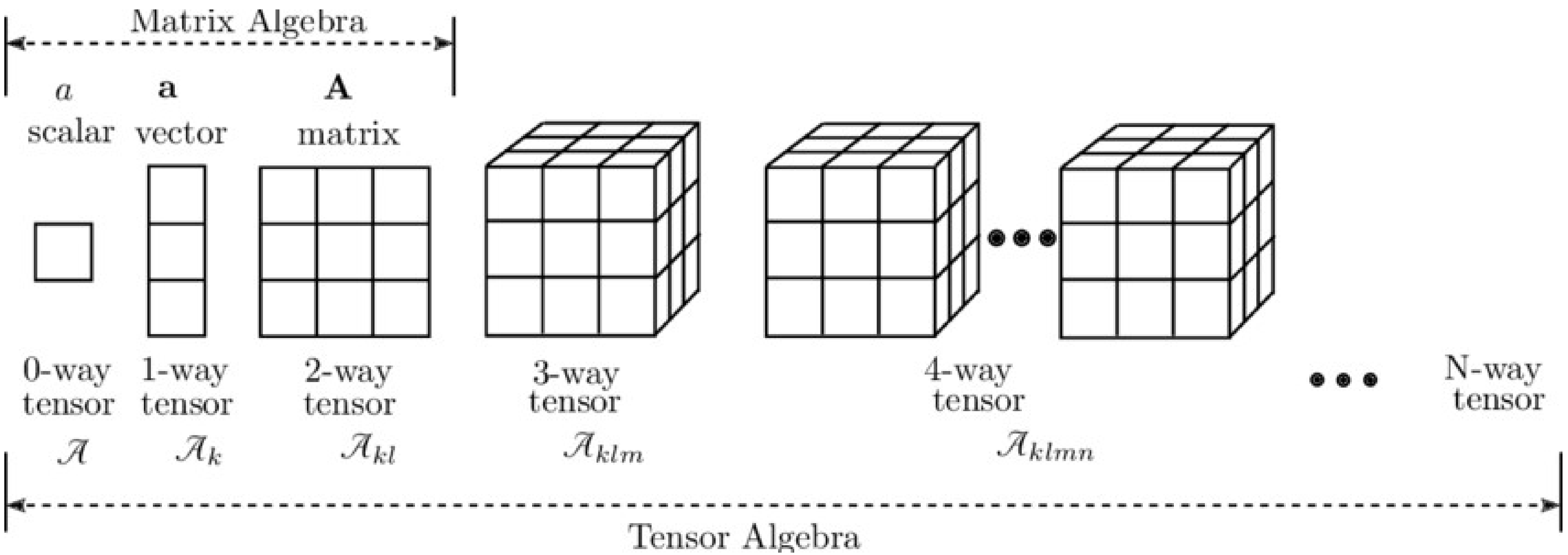
Introduction to AI for postgraduate students

Lecture Note 2-1
Linear Algebra

Hyun Jong Yang (hyunyang@postech.ac.kr)

Definition

- Scalar: single number
- Vector: array of numbers
- Matrix: 2-D array of numbers
- Tensors: array with more than two axes



Operations

- Transpose

$$\boldsymbol{A} = \begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \\ A_{3,1} & A_{3,2} \end{bmatrix} \Rightarrow \boldsymbol{A}^{\top} = \begin{bmatrix} A_{1,1} & A_{2,1} & A_{3,1} \\ A_{1,2} & A_{2,2} & A_{3,2} \end{bmatrix}$$

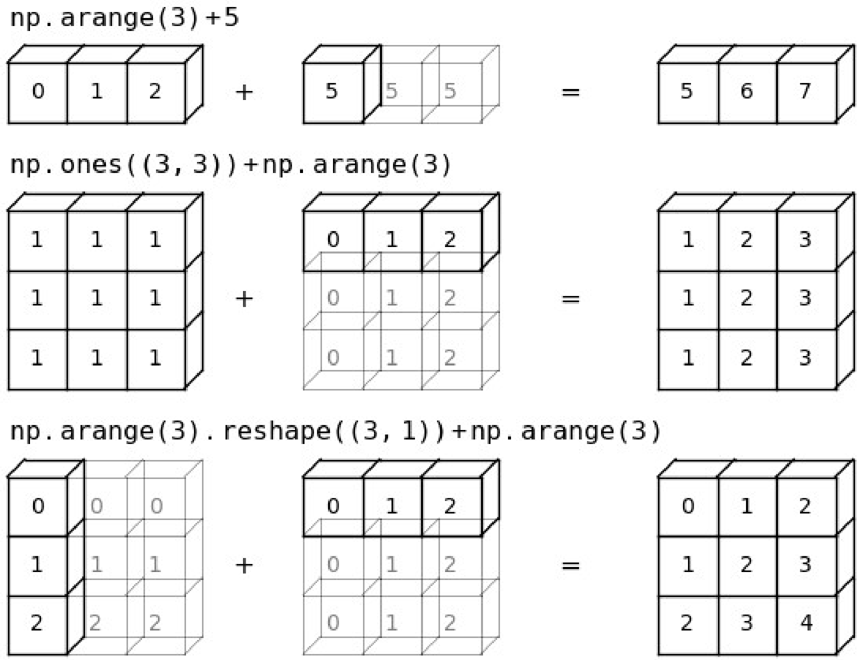
- Broadcasting

vector

$$\boldsymbol{C} = \boldsymbol{A} + \boldsymbol{b}$$

Matrix

$$C_{i,j} = A_{i,j} + b_j$$



Multiplying Matrices and Vectors

- Matrix multiplication:
  - Dimensions of $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$: $(m, n), (n, p), (m, p)$

$$\boldsymbol{C} = \boldsymbol{A}\boldsymbol{B}$$

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

  - Dot product: $\boldsymbol{x} \cdot \boldsymbol{y} = \boldsymbol{x}^{\mathrm{T}} \boldsymbol{y}$

- Hadamard product

$$\overset{G}{\begin{bmatrix} 3 & 5 & 7 \\ 4 & 9 & 8 \end{bmatrix}} \circ \overset{H}{\begin{bmatrix} 1 & 6 & 3 \\ 0 & 2 & 9 \end{bmatrix}} = \overset{N}{\begin{bmatrix} 3\times1 & 5\times6 & 7\times3 \\ 4\times0 & 9\times2 & 8\times9 \end{bmatrix}}$$

Multiplying Matrices and Vectors

- Properties of matrix multiplication
  - Distributive

$$\boldsymbol{A}(\boldsymbol{B}+\boldsymbol{C})=\boldsymbol{A}\boldsymbol{B}+\boldsymbol{A}\boldsymbol{C}.$$

  - Associative

$$\boldsymbol{A}(\boldsymbol{B}\boldsymbol{C})=(\boldsymbol{A}\boldsymbol{B})\boldsymbol{C}$$

  - Not commutative for matrix multiplication

$$\boldsymbol{A}\boldsymbol{B}\neq\boldsymbol{B}\boldsymbol{A}$$

  - But commutative for dot product

$$\boldsymbol{x}^\top\boldsymbol{y}=\boldsymbol{y}^\top\boldsymbol{x}$$

Identity and Inverse Matrices

- Identity matrix

$$\forall \boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{I}_n \boldsymbol{x} = \boldsymbol{x} \quad \boldsymbol{I}_n \in \mathbb{R}^{n \times n}$$

$$\boldsymbol{I}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

- Inverse matrix

$$\boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}_n$$

- Solving a linear equation

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$$

$$\boldsymbol{A}^{-1}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}$$

$$\boldsymbol{I}_n\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}$$

$$\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}.$$

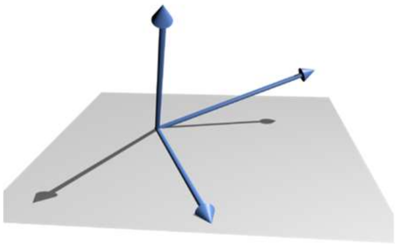
Linear Dependence

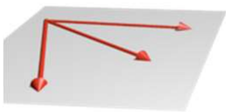
- Linearly dependent

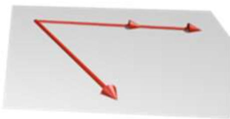
A sequence of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ from a vector space V is said to be *linearly dependent*, if there exist scalars $a_1, a_2, \ldots, a_k$, not all zero, such that

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_k\mathbf{v}_k = \mathbf{0},$$

where $\mathbf{0}$ denotes the zero vector.



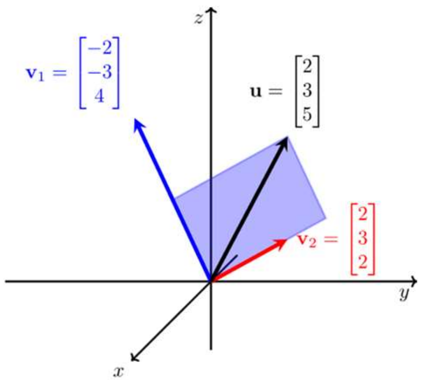




Span

- Span

$$\mathrm{span}(S) = \left\{ \sum_{i=1}^{k} \lambda_i v_i \;\middle|\; k \in \mathbb{N}, v_i \in S, \lambda_i \in K \right\}$$



Norms

- L^p norm for $p \in \mathbb{R}, p \geq 1$:

$$||\boldsymbol{x}||_p = \left(\sum_i |x_i|^p \right)^{\frac{1}{p}}$$

- L^2 norm: Euclidean norm
- L^1 norm: $||\boldsymbol{x}||_1 = \sum_i |x_i|$
- L^∞ norm: $||\boldsymbol{x}||_\infty = \max_i |x_i|$
- Frobenius norm: $||A||_F = \sqrt{\sum_{i,j} A_{i,j}^2}$

Special Kinds of Matrices and Vectors

- Diagonal matrix

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

- Symmetric matrix

$$\boldsymbol{A} = \boldsymbol{A}^\top$$

- Unit vector

$$||\boldsymbol{x}||_2 = 1$$

- Orthogonality

$$\boldsymbol{x}^\top \boldsymbol{y} = 0$$

- Orthogonal matrix

$$\boldsymbol{A}^\top \boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}^\top = \boldsymbol{I}$$

$$\boldsymbol{A}^{-1} = \boldsymbol{A}^\top$$

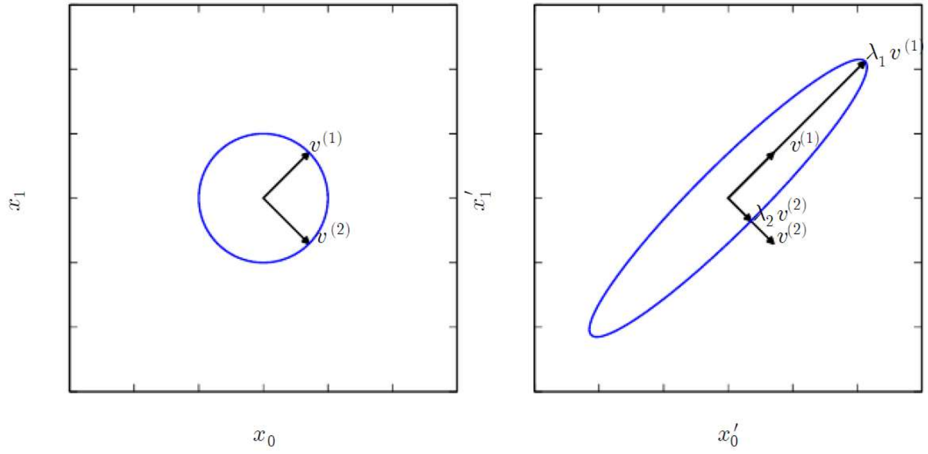
Eigen Decomposition

- Square matrix: A, Eigen vector: v, Eigen value: λ

$$\boldsymbol{A}\boldsymbol{v} = \lambda\boldsymbol{v}$$

$$\boldsymbol{A}[\boldsymbol{v}_1\ \boldsymbol{v}_2] = [\boldsymbol{v}_1\ \boldsymbol{v}_2]\begin{bmatrix}\lambda_1 & 0 \\ 0 & \lambda_2\end{bmatrix}$$

$$\boldsymbol{A} = \boldsymbol{V}\operatorname{diag}(\boldsymbol{\lambda})\boldsymbol{V}^{-1}$$



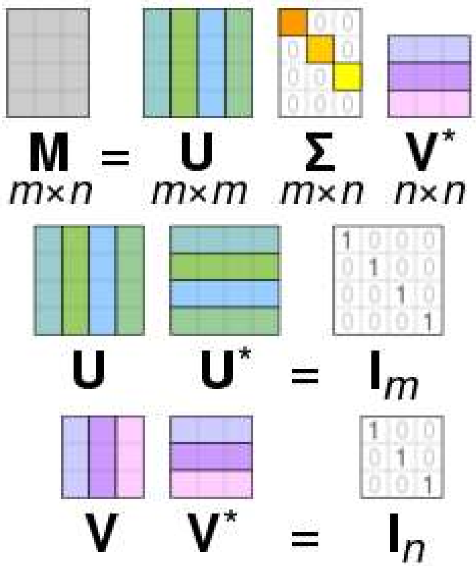
Singular Value Decomposition

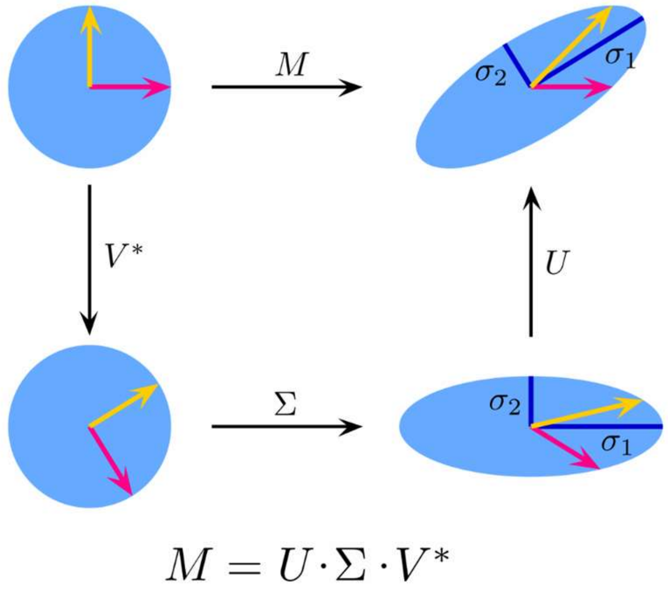
Singular values
(diagonal matrix)

$$A = UDV^{\top}$$

Left singular matrix
(orthogonal matrix)

Right singular matrix
(orthogonal matrix)





Moore-Penrose Pseudoinverse

- Moore-Penrose pseudoinverse matrix

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

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

  - $\boldsymbol{D}^{+}$of a diagonal matrix $\boldsymbol{D}$ is obtained by taking the reciprocal of its nonzero elements then taking the transpose of the resulting matrix.

Trace Operator

- Definition

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

- Frobenius norm

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

- Properties

$$\mathrm{Tr}(\boldsymbol{A}) = \mathrm{Tr}(\boldsymbol{A}^\top)$$

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

Determinant

- Definition

$$\det(A) = \prod_{i=1}^{n} \lambda_i = \lambda_1 \lambda_2 \cdots \lambda_n$$

- Examples

$$|A| = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc.$$

$$\begin{aligned} |A| = \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} &= a \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \begin{vmatrix} d & e \\ g & h \end{vmatrix} \\ &= aei + bfg + cdh - ceg - bdi - afh. \end{aligned}$$

Vector Calculus

- Gradient of f (with respect to $A \in \mathbb{R}^{m\times n}$)

$$\nabla_A f(A) \in \mathbb{R}^{m\times n} = \begin{bmatrix} \frac{\partial f(A)}{\partial A_{11}} & \frac{\partial f(A)}{\partial A_{12}} & \cdots & \frac{\partial f(A)}{\partial A_{1n}} \\ \frac{\partial f(A)}{\partial A_{21}} & \frac{\partial f(A)}{\partial A_{22}} & \cdots & \frac{\partial f(A)}{\partial A_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f(A)}{\partial A_{m1}} & \frac{\partial f(A)}{\partial A_{m2}} & \cdots & \frac{\partial f(A)}{\partial A_{mn}} \end{bmatrix}$$

- Linearity of Gradient

$$\nabla_x(f(x) + g(x)) = \nabla_x f(x) + \nabla_x g(x).$$

$$\text{For } t \in \mathbb{R},\ \nabla_x(t\ f(x)) = t\nabla_x f(x).$$

Vector Calculus

- Examples

$$f(x) = b^T x \quad \Rightarrow \quad f(x) = \sum_{i=1}^{n} b_i x_i \quad \Rightarrow \quad \frac{\partial f(x)}{\partial x_k} = \frac{\partial}{\partial x_k} \sum_{i=1}^{n} b_i x_i = b_k. \quad \Rightarrow \quad b$$

$$f(x) = x^T A x \quad \Rightarrow \quad f(x) = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} x_i x_j$$

$$\begin{aligned} \Rightarrow \quad \frac{\partial f(x)}{\partial x_k} &= \frac{\partial}{\partial x_k} \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} x_i x_j \\ &= \frac{\partial}{\partial x_k} \left[\sum_{i \neq k} \sum_{j \neq k} A_{ij} x_i x_j + \sum_{i \neq k} A_{ik} x_i x_k + \sum_{j \neq k} A_{kj} x_k x_j + A_{kk} x_k^2 \right] \\ &= \sum_{i \neq k} A_{ik} x_i + \sum_{j \neq k} A_{kj} x_j + 2 A_{kk} x_k \\ &= \sum_{i=1}^{n} A_{ik} x_i + \sum_{j=1}^{n} A_{kj} x_j = 2 \sum_{i=1}^{n} A_{ki} x_i, \quad \Rightarrow \quad 2Ax \end{aligned}$$

Vector Calculus

- Hessian

$$\nabla_x^2 f(x) \in \mathbb{R}^{n\times n} = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix}$$

- Hessian is symmetric since $\displaystyle \frac{\partial^2 f(x)}{\partial x_i \partial x_j} = \frac{\partial^2 f(x)}{\partial x_j \partial x_i}.$

- Example

$$f(x) = x^T A x \quad \Rightarrow \quad f(x) = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} x_i x_j$$

$$\Rightarrow \quad \frac{\partial^2 f(x)}{\partial x_k \partial x_\ell} = \frac{\partial}{\partial x_k}\left[\frac{\partial f(x)}{\partial x_\ell}\right] = \frac{\partial}{\partial x_k}\left[2\sum_{i=1}^{n} A_{\ell i} x_i\right] = 2A_{\ell k} = 2A_{k\ell}. \quad \Rightarrow \quad \nabla_x^2 x^T A x = 2A$$

Principal Components Analysis

- Original Data: $\{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(m)}\}$ in $\mathbb{R}^n$
- Goal: apply lossy compression to these points, which requires less memory losing as little precision as possible to get the mapped code vectors $\boldsymbol{c}^{(i)} \in \mathbb{R}^l$ ($\ell < n$)
- Assumption: *Linear* encoding/decoding

Encoding: $f(\boldsymbol{x}) = \boldsymbol{c}$

Encoding function

Decoding: $\boldsymbol{x} \approx g(f(\boldsymbol{x})) = \boldsymbol{D}\boldsymbol{c}$

Decoding function

- Part1: Optimizing the code *c* vector for given *x*

$$\boldsymbol{c}^* = \arg\min_{\boldsymbol{c}} ||\boldsymbol{x} - g(\boldsymbol{c})||_2.$$

$$= \arg\min_{\boldsymbol{c}} ||\boldsymbol{x} - g(\boldsymbol{c})||_2^2.$$

Principal Components Analysis

- Part1: Optimizing the code *c* vector for given *x*

$$
\begin{aligned}
(\boldsymbol{x} - g(\boldsymbol{c}))^\top(\boldsymbol{x} - g(\boldsymbol{c})) &= \boldsymbol{x}^\top\boldsymbol{x} - \boldsymbol{x}^\top g(\boldsymbol{c}) - g(\boldsymbol{c})^\top\boldsymbol{x} + g(\boldsymbol{c})^\top g(\boldsymbol{c}) \\
&= \boldsymbol{x}^\top\boldsymbol{x} \boxed{- 2\boldsymbol{x}^\top g(\boldsymbol{c}) + g(\boldsymbol{c})^\top g(\boldsymbol{c})}
\end{aligned}
$$

$$
\begin{aligned}
\boldsymbol{c}^* &= \arg\min_{\boldsymbol{c}} -2\boldsymbol{x}^\top g(\boldsymbol{c}) + g(\boldsymbol{c})^\top g(\boldsymbol{c}) \\
&= \arg\min_{\boldsymbol{c}} -2\boldsymbol{x}^\top\boldsymbol{D}\boldsymbol{c} + \boldsymbol{c}^\top\boldsymbol{D}^\top\boldsymbol{D}\boldsymbol{c} \\
&= \arg\min_{\boldsymbol{c}} -2\boldsymbol{x}^\top\boldsymbol{D}\boldsymbol{c} + \boldsymbol{c}^\top\boldsymbol{I}_l\boldsymbol{c} \\
&= \arg\min_{\boldsymbol{c}} -2\boldsymbol{x}^\top\boldsymbol{D}\boldsymbol{c} + \boldsymbol{c}^\top\boldsymbol{c}
\end{aligned}
$$

Principal Components Analysis

$$\nabla_{\boldsymbol{c}}(-2\boldsymbol{x}^\top \boldsymbol{D}\boldsymbol{c} + \boldsymbol{c}^\top \boldsymbol{c}) = \boldsymbol{0}$$

$$-2\boldsymbol{D}^\top \boldsymbol{x} + 2\boldsymbol{c} = \boldsymbol{0}$$ ➡ $$f(\boldsymbol{x}) = \boldsymbol{D}^\top \boldsymbol{x}$$ ➡ $$r(\boldsymbol{x}) = g\left(f\left(\boldsymbol{x}\right)\right) = \boldsymbol{D}\boldsymbol{D}^\top \boldsymbol{x}$$

$$\boldsymbol{c} = \boldsymbol{D}^\top \boldsymbol{x}.$$

- Part 2: Choosing the encoding matrix *D*

$$\boldsymbol{D}^* = \arg\min_{\boldsymbol{D}} \sqrt{\sum_{i,j}\left(x_j^{(i)} - r(\boldsymbol{x}^{(i)})_j\right)^2} \text{ subject to } \boldsymbol{D}^\top \boldsymbol{D} = \boldsymbol{I}_l$$

Assumption: ℓ=1 $D \rightarrow d$

➡
$$\boldsymbol{d}^* = \arg\min_{\boldsymbol{d}} \sum_i ||\boldsymbol{x}^{(i)} - \boldsymbol{d}\boldsymbol{d}^\top \boldsymbol{x}^{(i)}||_2^2 \text{ subject to } ||\boldsymbol{d}||_2 = 1$$
$$= \arg\min_{\boldsymbol{d}} \sum_i ||\boldsymbol{x}^{(i)} - \boldsymbol{d}^\top \boldsymbol{x}^{(i)} \boldsymbol{d}||_2^2 \text{ subject to } ||\boldsymbol{d}||_2 = 1$$
$$= \arg\min_{\boldsymbol{d}} \sum_i ||\boldsymbol{x}^{(i)} - \boldsymbol{x}^{(i)\top} \boldsymbol{d}\boldsymbol{d}||_2^2 \text{ subject to } ||\boldsymbol{d}||_2 = 1$$

Principal Components Analysis

Defining the augmented matrix $\boldsymbol{X}_{i,:} = \boldsymbol{x}^{(i)\top}$

$$\boldsymbol{d}^* = \arg\min_{\boldsymbol{d}} \sum_i ||\boldsymbol{x}^{(i)} - \boldsymbol{x}^{(i)\top}\boldsymbol{d}\boldsymbol{d}||_2^2 \text{ subject to } ||\boldsymbol{d}||_2 = 1$$

$$= \arg\min_{\boldsymbol{d}} ||\boldsymbol{X} - \boldsymbol{X}\boldsymbol{d}\boldsymbol{d}^\top||_F^2 \text{ subject to } \boldsymbol{d}^\top\boldsymbol{d} = 1$$

From

$$\arg\min_{\boldsymbol{d}} ||\boldsymbol{X} - \boldsymbol{X}\boldsymbol{d}\boldsymbol{d}^\top||_F^2 = \arg\min_{\boldsymbol{d}} \mathrm{Tr}\left(\left(\boldsymbol{X} - \boldsymbol{X}\boldsymbol{d}\boldsymbol{d}^\top\right)^\top \left(\boldsymbol{X} - \boldsymbol{X}\boldsymbol{d}\boldsymbol{d}^\top\right)\right)$$

$$= \arg\min_{\boldsymbol{d}} \mathrm{Tr}(\boldsymbol{X}^\top\boldsymbol{X} - \boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{d}\boldsymbol{d}^\top - \boldsymbol{d}\boldsymbol{d}^\top\boldsymbol{X}^\top\boldsymbol{X} + \boldsymbol{d}\boldsymbol{d}^\top\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{d}\boldsymbol{d}^\top)$$

$$= \arg\min_{\boldsymbol{d}} \mathrm{Tr}(\boldsymbol{X}^\top\boldsymbol{X}) - \mathrm{Tr}(\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{d}\boldsymbol{d}^\top) - \mathrm{Tr}(\boldsymbol{d}\boldsymbol{d}^\top\boldsymbol{X}^\top\boldsymbol{X}) + \mathrm{Tr}(\boldsymbol{d}\boldsymbol{d}^\top\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{d}\boldsymbol{d}^\top)$$

$$= \arg\min_{\boldsymbol{d}} - \mathrm{Tr}(\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{d}\boldsymbol{d}^\top) - \mathrm{Tr}(\boldsymbol{d}\boldsymbol{d}^\top\boldsymbol{X}^\top\boldsymbol{X}) + \mathrm{Tr}(\boldsymbol{d}\boldsymbol{d}^\top\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{d}\boldsymbol{d}^\top)$$

$$= \arg\min_{\boldsymbol{d}} -2\,\mathrm{Tr}(\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{d}\boldsymbol{d}^\top) + \mathrm{Tr}(\boldsymbol{d}\boldsymbol{d}^\top\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{d}\boldsymbol{d}^\top)$$

$$= \arg\min_{\boldsymbol{d}} -2\,\mathrm{Tr}(\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{d}\boldsymbol{d}^\top) + \mathrm{Tr}(\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{d}\boldsymbol{d}^\top\boldsymbol{d}\boldsymbol{d}^\top)$$

Principal Components Analysis

Thus, we have

$$\arg\min_{\boldsymbol{d}} -2\,\mathrm{Tr}(\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{d}\boldsymbol{d}^\top) + \mathrm{Tr}(\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{d}\boldsymbol{d}^\top \boldsymbol{d}\boldsymbol{d}^\top) \text{ subject to } \boldsymbol{d}^\top \boldsymbol{d} = 1$$

$$= \arg\min_{\boldsymbol{d}} -2\,\mathrm{Tr}(\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{d}\boldsymbol{d}^\top) + \mathrm{Tr}(\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{d}\boldsymbol{d}^\top) \text{ subject to } \boldsymbol{d}^\top \boldsymbol{d} = 1$$

$$= \arg\min_{\boldsymbol{d}} -\mathrm{Tr}(\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{d}\boldsymbol{d}^\top) \text{ subject to } \boldsymbol{d}^\top \boldsymbol{d} = 1$$

$$= \arg\max_{\boldsymbol{d}} \mathrm{Tr}(\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{d}\boldsymbol{d}^\top) \text{ subject to } \boldsymbol{d}^\top \boldsymbol{d} = 1$$

$$= \arg\max_{\boldsymbol{d}} \mathrm{Tr}(\boldsymbol{d}^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{d}) \text{ subject to } \boldsymbol{d}^\top \boldsymbol{d} = 1.$$

$||\boldsymbol{X}\boldsymbol{d}||^2$

Principal Components Analysis

Therefore, the optimal

$$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\boldsymbol{T}}$$

$$\begin{aligned}
\boldsymbol{d}^* = \arg\max ||\boldsymbol{X}\boldsymbol{d}||^2 &= \arg\max \left\|\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\boldsymbol{T}}\boldsymbol{d}\right\|^{\boldsymbol{2}} \\
&= \arg\max \left(\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{T}\boldsymbol{d}\right)^{T}\left(\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\boldsymbol{T}}\boldsymbol{d}\right) \\
&= \arg\max \boldsymbol{d}^{T}\boldsymbol{V}\boldsymbol{S}^{\boldsymbol{2}}\boldsymbol{V}^{\boldsymbol{T}}\,\boldsymbol{d}
\end{aligned}$$

➔ the optimal $\boldsymbol{d}$ is the right singular vector corresponding to the largest singular value of $\boldsymbol{X}$