12. Principle Components Analysis I

$$\underset{n \times m}{\mathbb{X}} = \{ X^{(1)}\ X^{(2)} \cdots X^{(m)} \} = \left\{ \begin{bmatrix} x_1^{(1)} \\ x_2^{(1)} \\ \vdots \\ x_n^{(1)} \end{bmatrix} \begin{bmatrix} x_1^{(2)} \\ x_2^{(2)} \\ \vdots \\ x_n^{(2)} \end{bmatrix} \cdots \begin{bmatrix} x_1^{(m)} \\ x_2^{(m)} \\ \vdots \\ x_n^{(m)} \end{bmatrix} \right\}$$

$$C^{(i)} = \begin{bmatrix} c_1^{(i)} \\ c_2^{(i)} \\ \vdots \\ c_l^{(i)} \end{bmatrix}, \quad l \leq n$$

$$\underset{n \times 1}{X^{(i)}} \xrightarrow{f} \underset{l \times 1}{C^{(i)}} \xrightarrow{g} \underset{n \times 1}{r^{(i)}} \Rightarrow X \cong g(f(X))$$

Let $g(C) = \underset{n \times l}{D} \underset{l \times 1}{C}$, where

$$D = \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1l} \\ d_{21} & d_{22} & \cdots & d_{2l} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n1} & d_{n2} & \cdots & d_{nl} \end{bmatrix} = [\, d^1\ d^2 \cdots d^l \,], \text{ where}$$

$(d^i)^T(d^j) = 0,\ i \neq j$ (orthogonal).

$\| d^i \|_2 = 1$ (unit norm).

$$C^* = \arg\min_C \| X - g(C) \|_2$$

$$= \arg\min_C \| X - g(C) \|_2^2$$

$$\| X - g(C) \|_2^2$$

$$= (X - g(C))^T (X - g(C))$$

$$= (X^T - g(C)^T)(X - g(C))$$

$$= X^T X - X^T g(C) - g(C)^T X + g(C)^T g(C)$$

$$= X^T X - 2 X^T g(C) + g(C)^T g(C)$$

$$c^* = \underset{C}{\arg\min}\ -2X^T g(C) + g(C)^T g(C)$$

$$= \underset{C}{\arg\min}\ -2X^T D C + (DC)^T DC$$

$$= \underset{C}{\arg\min}\ -2X^T D C + C^T I_{\ell\times\ell}\, C$$

$$= \underset{C}{\arg\min}\ -2X^T D C + C^T C$$

$$-2X^T D C + C^T C$$

$$= -2 \cdot \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \cdot \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1\ell} \\ d_{21} & d_{22} & \cdots & d_{2\ell} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n1} & d_{n2} & \cdots & d_{n\ell} \end{bmatrix} \cdot \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_\ell \end{bmatrix} + \begin{bmatrix} c_1 & c_2 & \cdots & c_\ell \end{bmatrix} \cdot \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_\ell \end{bmatrix}$$

$$= -2 \cdot \begin{bmatrix} x_1 d_{11} + \cdots + x_n d_{n1} & \circ\circ\circ & x_1 d_{1\ell} + \cdots + x_n d_{n\ell} \end{bmatrix} \cdot \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_\ell \end{bmatrix} + \sum_{i=1}^{\ell} c_i^2$$

$$= -2 \cdot \left[(x_1 d_{11} + \cdots + x_n d_{n1}) c_1 + \circ\circ\circ + (x_1 d_{1\ell} + \cdots + x_n d_{n\ell}) c_\ell \right] + \sum_{i=1}^{\ell} c_i^2$$

$$\nabla_c (-2X^T D C + C^T c)$$

$$= \begin{bmatrix} \frac{\partial}{\partial c_1} \left(-2 \left[(x_1 d_{11} + \cdots + x_n d_{n1}) c_1 + \circ\circ\circ + (x_1 d_{1\ell} + \cdots + x_n d_{n\ell}) c_\ell \right] + \sum_{i=1}^{\ell} c_i^2 \right) \\ \circ \\ \circ \\ \circ \\ \frac{\partial}{\partial c_\ell} \left(-2 \left[(x_1 d_{11} + \cdots + x_n d_{n1}) c_1 + \circ\circ\circ + (x_1 d_{1\ell} + \cdots + x_n d_{n\ell}) c_\ell \right] + \sum_{i=1}^{\ell} c_i^2 \right) \end{bmatrix}$$

$$= \begin{bmatrix} -2(x_1 d_{11} + \cdots + x_n d_{n1}) + 2c_1 \\ \circ \\ \circ \\ \circ \\ -2(x_1 d_{1\ell} + \cdots + x_n d_{n\ell}) + 2c_\ell \end{bmatrix}$$

$$= -2 \begin{bmatrix} x_1 d_{11} + \cdots + x_n d_{n1} \\ \vdots \\ x_1 d_{1\ell} + \cdots + x_n d_{n\ell} \end{bmatrix} + 2 \begin{bmatrix} c_1 \\ \vdots \\ c_\ell \end{bmatrix}$$

$$= -2 \begin{bmatrix} d_{11} & d_{21} & \cdots & d_{n1} \\ d_{12} & d_{22} & \cdots & d_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ d_{1\ell} & d_{2\ell} & \cdots & d_{n\ell} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + 2 \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_\ell \end{bmatrix}$$

$$= -2\, D^T X + 2\, C = 0$$

Note on Differentiation w.r.t vector: $\underset{n\times 1}{X}, \underset{m\times 1}{Y}, \underset{n\times 1}{W}, \underset{m\times n}{A}, \underset{n\times m}{B}$

$f_1(X) = X^T W,\ f_2(X) = X^T X,\ f_3(X) = X^T B Y,\ f_4(X) = Y^T A X$

① $f_1'(X) = \nabla_X (X^T W) = W$

② $f_2'(X) = \nabla_X (X^T X) = 2X$

③ $f_4'(X) = \nabla_X (Y^T A X) = \nabla_X (X^T (Y^T A)^T) = \nabla_X (X^T A^T Y) \overset{\text{by ①}}{=} A^T Y$

④ $f_3'(X) = \nabla_X (X^T B Y) = \nabla_X (Y^T (X^T B)^T) = \nabla_X (Y^T B^T X) \underset{\text{by ③}}{=} (B^T)^T Y = BY$

$\therefore\ C^* = D^T X \quad$ s.t. $\| X - g(C^*) \|_2 = \min\limits_{C} \| X - g(C) \|_2$

Let $f(X) = D^T X$, then

$$r(X) = g(f(X)) = g(D^T X) = D D^T X$$

$$D^* = \arg\min_D \Big(\sum_{i=1}^{m} \sum_{j=1}^{n} \big(X_j^{(i)} - r(X^{(i)})_j \big)^2 \Big)^{\frac{1}{2}} \quad \text{subject to } D^T D = I_{\ell\times\ell}$$

$$= \arg\min_D \Big(\sum_{i=1}^{m} \sum_{j=1}^{n} \big(X_j^{(i)} - (D D^T X^{(i)})_j \big)^2 \Big)^{\frac{1}{2}}, \quad D^T D = I_{\ell\times\ell}$$

$$= \arg\min_D \sum_{i=1}^{m} \| X^{(i)} - D D^T X^{(i)} \|_2^2, \quad D^T D = I_{\ell\times\ell}.$$

Let $\ell = 1$, then $D = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{bmatrix} = d$, and

$$d^* = \arg\min_d \sum_{i=1}^{m} \| x^{(i)} - \underset{n\times1}{d}\,\underbrace{\underset{1\times n}{d^T}\,\underset{n\times1}{x^{(i)}}}_{\text{scalar}} \|_2^2 , \quad d^T d = 1$$

$$= \arg\min_d \sum_{i=1}^{m} \| x^{(i)} - \underbrace{d^T x^{(i)}}_{\text{scalar}}\, d \|_2^2 , \quad \|d\|_2 = 1$$

$$= \arg\min_d \sum_{i=1}^{m} \| x^{(i)} - (d^T x^{(i)})^T d \|_2^2 , \quad \|d\|_2 = 1$$

$$= \arg\min_d \sum_{i=1}^{m} \| x^{(i)} - (x^{(i)})^T d\, d \|_2^2 , \quad \|d\|_2 = 1$$

$$I = \sum_{i=1}^{m} \| x^{(i)} - (x^{(i)})^T d\, d \|_2^2$$

$$= \sum_{i=1}^{m} \| \begin{bmatrix} x_1^{(i)} \\ x_2^{(i)} \\ \vdots \\ x_n^{(i)} \end{bmatrix} - \begin{bmatrix} x_1^{(i)} & x_2^{(i)} & \cdots & x_n^{(i)} \end{bmatrix} \cdot \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{bmatrix} \|_2^2$$

$$\alpha^{(i)} = x_1^{(i)} d_1 + x_2^{(i)} d_2 + \cdots + x_n^{(i)} d_n .$$

$$I = \sum_{i=1}^{m} \| \begin{bmatrix} x_1^{(i)} - \alpha^{(i)} \cdot d_1 \\ x_2^{(i)} - \alpha^{(i)} \cdot d_2 \\ \vdots \\ x_n^{(i)} - \alpha^{(i)} \cdot d_n \end{bmatrix} \|_2^2$$

$$= \sum_{i=1}^{m} (x_1^{(i)} - \alpha^{(i)} d_1)^2 + (x_2^{(i)} - \alpha^{(i)} d_2)^2 + \cdots + (x_n^{(i)} - \alpha^{(i)} d_n)^2$$

$$= \| \begin{bmatrix} x_1^{(1)} - \alpha^{(1)} d_1 & x_2^{(1)} - \alpha^{(1)} d_2 & \cdots & x_n^{(1)} - \alpha^{(1)} d_n \\ x_1^{(2)} - \alpha^{(2)} d_1 & x_2^{(2)} - \alpha^{(2)} d_2 & \cdots & x_n^{(2)} - \alpha^{(2)} d_n \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{(m)} - \alpha^{(m)} d_1 & x_2^{(m)} - \alpha^{(m)} d_2 & \cdots & x_n^{(m)} - \alpha^{(m)} d_n \end{bmatrix} \|_2^2$$

$$= \left\| \begin{bmatrix} x_1^{(1)} & x_2^{(1)} & \cdots & x_n^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \cdots & x_n^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{(m)} & x_2^{(m)} & \cdots & x_n^{(m)} \end{bmatrix} - \begin{bmatrix} \alpha^{(1)} \\ \alpha^{(2)} \\ \vdots \\ \alpha^{(m)} \end{bmatrix} \cdot \begin{bmatrix} d_1 & d_2 & \cdots & d_n \end{bmatrix} \right\|_F^2$$

$$= \left\| \begin{bmatrix} x_1^{(1)} & x_2^{(1)} & \cdots & x_n^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \cdots & x_n^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{(m)} & x_2^{(m)} & \cdots & x_n^{(m)} \end{bmatrix} - \begin{bmatrix} x_1^{(1)} & x_2^{(1)} & \cdots & x_n^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \cdots & x_n^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{(m)} & x_2^{(m)} & \cdots & x_n^{(m)} \end{bmatrix} \cdot \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{bmatrix} \cdot \begin{bmatrix} d_1 & d_2 & \cdots & d_n \end{bmatrix} \right\|_F^2$$

$= \| A - A d d^T \|_F^2$, where

$$A = \mathbb{X}^T = \begin{bmatrix} [x_1^{(1)} & x_2^{(1)} & \cdots & x_n^{(1)}] \\ [x_1^{(2)} & x_2^{(2)} & \cdots & x_n^{(2)}] \\ \vdots & \vdots & \ddots & \vdots \\ [x_1^{(m)} & x_2^{(m)} & \cdots & x_n^{(m)}] \end{bmatrix}$$

$$\therefore\ d^* = \arg\min_d \| A - A d d^T \|_F^2, \quad \| d \|_2 = 1$$

$$= \arg\min_d \left(\left(Tr\left((A - A d d^T)(A - A d d^T)^T \right) \right)^{\frac{1}{2}} \right)^2$$

$$= \arg\min_d Tr\left((A - A d d^T)^T (A - A d d^T) \right)$$

$$= \arg\min_d Tr\left((A^T - (A d d^T)^T)(A - A d d^T) \right)$$

$$= \arg\min_d Tr\left((A^T - d d^T A^T)(A - A d d^T) \right)$$

$$= \arg\min_{d} \mathrm{Tr}(A^TA - A^TAdd^T - dd^TA^TA + dd^TA^TAdd^T)$$

$$= \arg\min_{d} \mathrm{Tr}(A^TA) - \mathrm{Tr}(A^TAdd^T) - \mathrm{Tr}(dd^TA^TA) + \mathrm{Tr}(dd^TA^TAdd^T)$$

$$= \arg\min_{d} -\mathrm{Tr}(A^TAdd^T) - \mathrm{Tr}(dd^TA^TA) + \mathrm{Tr}(dd^TA^TA\,dd^T)$$

$$= \arg\min_{d} -2\,\mathrm{Tr}(A^TA\,dd^T) + \mathrm{Tr}(A^TA\,dd^Tdd^T)$$

$$= \arg\min_{d} -2\,\mathrm{Tr}(A^TA\,dd^T) + \mathrm{Tr}(A^TA\,dd^T)$$

$$= \arg\min_{d} -\mathrm{Tr}(A^TA\,dd^T)$$

$$= \arg\max_{d} \mathrm{Tr}(A^TA\,dd^T)$$

$$= \arg\max_{d} \mathrm{Tr}(\underbrace{d^TA^TA\,d}_{\text{scalar}})$$

$$= \arg\max_{d} d^T\underbrace{A^TA}_{\text{symmetric, } n\times n}d \quad \text{subject to } \|d\|_2 = 1$$

$$= \arg\max_{d} \{\lambda : A^TA \cdot d = \lambda d \wedge \|d\|_2 = 1\}$$

(Eigendecomposition of symmetric matrix A^TA)

In general,

$D = \{d_1, d_2, \cdots, d_\ell\}$ s.t.

1) $\|d_i\|_2 = 1, \quad i = 1, 2, \cdots, \ell$

2) $A^TA\,d_i = \lambda_i d_i, \ \lambda_i = d_i^TA^TA\,d_i, \quad i = 1, 2, \cdots, \ell$

3) $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{\ell-1} \geq \lambda_\ell \geq \lambda_{\ell+1} \geq \cdots$

6. Principal Components Analysis II

$$\underset{m\times n}{\mathbb{X}} = \begin{bmatrix} x^{(1)} \\ x^{(2)} \\ \vdots \\ x^{(m)} \end{bmatrix} = \begin{bmatrix} [x_1^{(1)} & x_2^{(1)} & \cdots & x_n^{(1)}] \\ [x_1^{(2)} & x_2^{(2)} & \cdots & x_n^{(2)}] \\ \vdots & \vdots & \ddots & \vdots \\ [x_1^{(m)} & x_2^{(m)} & \cdots & x_n^{(m)}] \end{bmatrix}$$

$$\hat{E}[\mathbb{X}] = \begin{bmatrix} \hat{E}[x_1] \\ \hat{E}[x_2] \\ \vdots \\ \hat{E}[x_n] \end{bmatrix} = \begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \vdots \\ \hat{u}_n \end{bmatrix} = \begin{bmatrix} \frac{1}{m}\cdot\sum_{j=1}^{m} x_1^{(j)} \\ \frac{1}{m}\cdot\sum_{j=1}^{m} x_2^{(j)} \\ \vdots \\ \frac{1}{m}\cdot\sum_{j=1}^{m} x_n^{(j)} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = 0 \quad \text{(Sample Mean Vector)}$$

$$\hat{Var}(x_i) = \hat{\sigma}_i^2 = \frac{1}{m-1}\cdot\sum_{j=1}^{m}(x_i^{(j)} - \hat{u}_i)^2$$

$$= \frac{1}{m-1}\cdot\sum_{j=1}^{m}(x_i^{(j)})^2, \quad i = 1, 2, 3, \cdots, n \quad \text{(Sample Variance)}$$

$\text{bias}(\hat{\sigma}_i^2) = 0$ (D.L., P 82)

$$\hat{Cov}(x_i, x_k) = \frac{1}{m-1}\cdot\sum_{j=1}^{m}(x_i^{(j)} - \hat{u}_i)(x_k^{(j)} - \hat{u}_k)$$

$$= \frac{1}{m-1}\cdot\sum_{j=1}^{m} x_i^{(j)}\cdot x_k^{(j)} \quad \text{(Sample Covariance)}$$

$$\hat{cov}(\mathbb{X}) = \begin{bmatrix} \hat{Var}(x_1) & \hat{Cov}(x_1, x_2) & \cdots & \hat{Cov}(x_1, x_n) \\ \hat{Cov}(x_2, x_1) & \hat{Var}(x_2) & \cdots & \hat{Cov}(x_2, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{Cov}(x_n, x_1) & \hat{Cov}(x_n, x_2) & \cdots & \hat{Var}(x_n) \end{bmatrix} \quad \text{(Sample Covariance Matrix)}$$

$$= \begin{bmatrix} \frac{1}{m-1}\cdot\sum_{j=1}^{m}(x_1^{(j)})^2 & \frac{1}{m-1}\cdot\sum_{j=1}^{m} x_1^{(j)}x_2^{(j)} & \cdots & \frac{1}{m-1}\cdot\sum_{j=1}^{m} x_1^{(j)}x_n^{(j)} \\ \frac{1}{m-1}\cdot\sum_{j=1}^{m} x_2^{(j)}x_1^{(j)} & \frac{1}{m-1}\cdot\sum_{j=1}^{m}(x_2^{(j)})^2 & \cdots & \frac{1}{m-1}\cdot\sum_{j=1}^{m} x_2^{(j)}x_n^{(j)} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{m-1}\cdot\sum_{j=1}^{m} x_n^{(j)}x_1^{(j)} & \frac{1}{m-1}\cdot\sum_{j=1}^{m} x_n^{(j)}x_2^{(j)} & \cdots & \frac{1}{m-1}\cdot\sum_{j=1}^{m}(x_n^{(j)})^2 \end{bmatrix}$$

$$= \frac{1}{m-1} \cdot \begin{bmatrix} \sum_{j=1}^{m} (x_1^{(j)})^2 & \sum_{j=1}^{m} x_1^{(j)} x_2^{(j)} & \cdots & \sum_{j=1}^{m} x_1^{(j)} x_n^{(j)} \\ \sum_{j=1}^{m} x_2^{(j)} x_1^{(j)} & \sum_{j=1}^{m} (x_2^{(j)})^2 & \cdots & \sum_{j=1}^{m} x_2^{(j)} x_n^{(j)} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{j=1}^{m} x_n^{(j)} x_1^{(j)} & \sum_{j=1}^{m} x_n^{(j)} x_2^{(j)} & \cdots & \sum_{j=1}^{m} (x_n^{(j)})^2 \end{bmatrix}$$

$$= \frac{1}{m-1} \cdot \begin{bmatrix} x_1^{(1)} & x_1^{(2)} & \cdots & x_1^{(m)} \\ x_2^{(1)} & x_2^{(2)} & \cdots & x_2^{(m)} \\ \vdots & \vdots & \ddots & \vdots \\ x_n^{(1)} & x_n^{(2)} & \cdots & x_n^{(m)} \end{bmatrix} \cdot \begin{bmatrix} x_1^{(1)} & x_2^{(1)} & \cdots & x_n^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \cdots & x_n^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{(m)} & x_2^{(m)} & \cdots & x_n^{(m)} \end{bmatrix}$$

$$= \frac{1}{m-1} \cdot \mathbb{X}^T \mathbb{X} \quad ①$$

If $\underset{n\times \ell}{D} = \{ d_1, d_2, \cdots, d_\ell \}$ s.t.

$\underset{n\times m}{\mathbb{X}^T} \underset{m\times n}{\mathbb{X}} \cdot \underset{n\times 1}{d_i} = \lambda_i \cdot d_i$ and $\| d_i \|_2 = 1$, then

$$\underset{1\times \ell}{\mathbf{Z}} = f(\mathbf{X}) = \underset{1\times n}{\mathbf{X}^T} \cdot \underset{n\times \ell}{D} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \cdot \begin{bmatrix} \begin{bmatrix} d_{11} \\ d_{21} \\ \vdots \\ d_{n1} \end{bmatrix} & \begin{bmatrix} d_{12} \\ d_{22} \\ \vdots \\ d_{n2} \end{bmatrix} & \cdots & \begin{bmatrix} d_{1\ell} \\ d_{2\ell} \\ \vdots \\ d_{n\ell} \end{bmatrix} \end{bmatrix}$$

$= \begin{bmatrix} z_1 & z_2 & \cdots & z_\ell \end{bmatrix}$, and

$$\underset{m\times \ell}{\mathbb{Z}} = f(\mathbb{X}) = \underset{m\times n}{\mathbb{X}} \cdot \underset{n\times \ell}{D} \quad ②$$

$$= \begin{bmatrix} \mathbf{X}^{(1)} \\ \mathbf{X}^{(2)} \\ \vdots \\ \mathbf{X}^{(m)} \end{bmatrix} \cdot \begin{bmatrix} d_1 & d_2 & \cdots & d_\ell \end{bmatrix}$$

$$= \begin{bmatrix} [x_1^{(1)} & x_2^{(1)} & \cdots & x_n^{(1)}] \\ [x_1^{(2)} & x_2^{(2)} & \cdots & x_n^{(2)}] \\ & \vdots & & \\ [x_1^{(m)} & x_2^{(m)} & \cdots & x_n^{(m)}] \end{bmatrix} \cdot \begin{bmatrix} \begin{bmatrix} d_{11} \\ d_{21} \\ \vdots \\ d_{n1} \end{bmatrix} & \begin{bmatrix} d_{12} \\ d_{22} \\ \vdots \\ d_{n2} \end{bmatrix} & \cdots & \begin{bmatrix} d_{1l} \\ d_{2l} \\ \vdots \\ d_{nl} \end{bmatrix} \end{bmatrix}$$

$$= \begin{bmatrix} z^{(1)} \\ z^{(2)} \\ \vdots \\ z^{(m)} \end{bmatrix}$$ (Principal Components Analysis I, D.L., P39)

$$\underset{m \times n}{\mathbb{X}} = \underset{m \times m}{U} \underset{m \times l}{\Sigma} \underset{l \times n}{W^T}$$ (Singular Value Decomposition, D.L., P 9),

Since $\mathbb{X}^T \mathbb{X}$ is symmetric,

$W = D$ ③ (D.L., P9), and

$\mathbb{X} = U \Sigma D^T$ ④, then

$\hat{Cov}(\mathbb{X})$

$$= \frac{1}{m-1} \cdot \mathbb{X}^T \mathbb{X}$$ (by ①)

$$= \frac{1}{m-1} \cdot (U \Sigma D^T)^T (U \Sigma D^T)$$ (by ④)

$$= \frac{1}{m-1} \cdot D \Sigma^T U^T U \Sigma D^T$$

$$= \frac{1}{m-1} \cdot D \Sigma^T \Sigma D^T$$ (U is orthogonal), and

$$\mathbb{X}^T \mathbb{X} = D \Sigma^T \Sigma D^T \quad ⑤$$

$\widehat{\text{Cov}}(\mathbb{Z})$

$$= \frac{1}{m-1} \cdot \mathbb{Z}^T \mathbb{Z}$$

$$= \frac{1}{m-1} \cdot (\mathbb{X} \cdot D)^T (\mathbb{X} \cdot D) \quad (\text{by } ②)$$

$$= \frac{1}{m-1} \cdot D^T \cdot \mathbb{X}^T \cdot \mathbb{X} \cdot D$$

$$= \frac{1}{m-1} \cdot D^T (D \Sigma^T \Sigma D^T) D \quad (\text{by } ⑤)$$

$$= \frac{1}{m-1} \cdot W^T W \ \Sigma^T \Sigma \ W^T W \quad (\text{by } ③)$$

$$= \frac{1}{m-1} \cdot \Sigma^T \Sigma \quad (W \text{ is orthogonal})$$

Since Σ is diagonal, and $\Sigma^T \Sigma$ is also diagonal, and

$\widehat{\text{Cov}}(z_i, z_j) = 0, \ i \neq j$.