Tutorial 7

7.4 - PCA w/ SVD

CS5487 Lecture Notes (2020)
Dr. Antoni B. Chan
Dept of Computer Science
City University of Hong Kong

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\Sigma = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)^T$$

a) $\bar{X} = [x_1-\mu, x_2-\mu, \ldots, x_n-\mu]$

$$= [x_1 \cdots x_n] - \underbrace{[\mu \cdots\cdots \mu]}_{n}$$

$$\mu 1^T = \begin{bmatrix} | \\ \mu \\ | \end{bmatrix} [- 1 -]$$

$$= X - \mu 1^T$$

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}[x_1 \cdots x_n]\begin{bmatrix}1\\ \vdots \\ 1\end{bmatrix} = \frac{1}{n} X 1$$

$$= X - \frac{1}{n} X 1 1^T$$

$$\boxed{\bar{X} = X(I - \tfrac{1}{n} 1 1^T)}$$ ← mean-subtracted points.

b) Singular Value Decomposition (SVD)

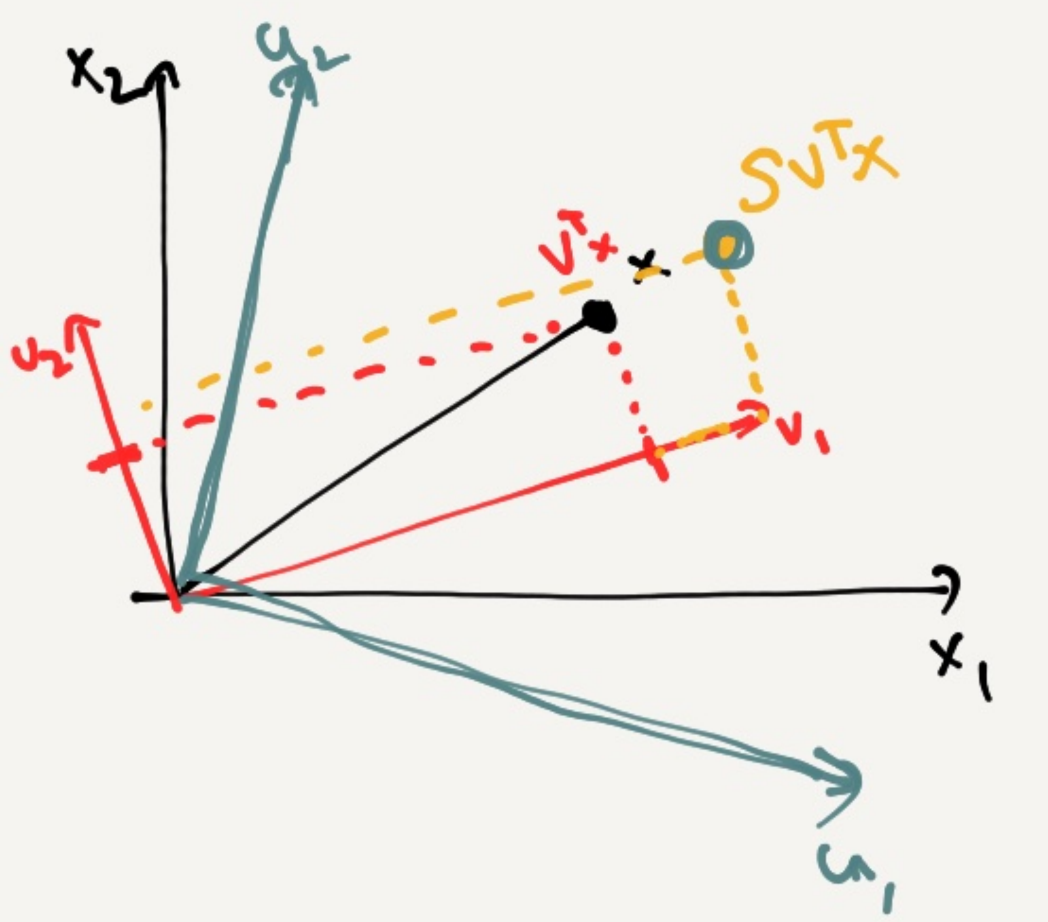
$$\underset{(n\times m)}{A} = \underset{n\times m}{U}\,\underset{m\times m}{S}\,\underset{m\times m}{V^T}$$

$$\begin{cases} U \in \mathbb{R}^{n\times m}, & U^T U = I \\ V \in \mathbb{R}^{m\times m}, & V^T V = I \\ S = \mathrm{diag}(s_1, \ldots, s_m) & s_i \geq 0 \quad (m\times m) \end{cases}$$

singular values

$$Ax = USV^T x$$

~ rotation
~ scaling
~ rotation



b) take SVD of $\bar{X}$:

$$\bar{X} = USV^T$$

$$\bar{X}^T = VSU^T$$

Covariance: $\Sigma = \frac{1}{n}\sum_i \underbrace{(x_i-\mu)}_{\bar{x}_i}\underbrace{(x_i-\mu)^T}_{\bar{x}_i^T} = \frac{1}{n}\bar{X}\bar{X}^T$

$$= \frac{1}{n}(US\underbrace{V^T)(V}_{I}SU^T)$$

$$= \frac{1}{n}US^2U^T = \underbrace{U\Big(\underbrace{\tfrac{1}{n}S^2}_{\text{eigenvalues}}\Big)\underbrace{U^T}_{\text{eigenvectors}}}$$

Thus: $(u_i, \frac{1}{n}s_i^2)$ is an eigen pair.

u_i: i^{th} column of U.

c) PCA using SVD

1) mean-subtract : $\bar{X} = X(I - \frac{1}{n}\mathbb{1}\mathbb{1}^T)$

2) SVD : $\bar{X} = USV^T$

3) sort by singular values: $s_1 > s_2 > \ldots > 0$

4) PCA matrix $\Phi = [u_1 \cdots u_K]$

5) PCA coefficient $z = \Phi^T(x-\mu)$

Problem 7.6

FLD

$$S_B = (\mu_1-\mu_2)(\mu_1-\mu_2)^T$$

$$S_W = S_1 + S_2$$

$$w^* = \underset{w}{\text{argmax}} \frac{w^TS_Bw}{w^TS_Ww}$$

← optimize

← fixed to 1

$$= \text{argmax } w^TS_Bw \quad \text{s.t. } w^TS_Ww = 1$$

Lagrangian:

$$L(w,\lambda) = w^TS_Bw - \lambda(w^TS_Ww - 1)$$

b) $\frac{\partial L}{\partial w} = \cancel{2}S_Bw - \lambda\cancel{2}S_Ww = 0$

$$\Rightarrow \boxed{S_Bw = \lambda \underbrace{S_W}w}$$

$S_W = \lambda w$

generalized eigenvalue problem

c) assume S_W is invertible, premultiply by S_W^{-1}

$$S_W^{-1}S_Bw = \lambda w$$

$$\underbrace{S_W^{-1}(\mu_1-\mu_2)}_{\text{vector}}\underbrace{(\mu_1-\mu_2)^Tw}_{\text{scalar}} = \underbrace{\lambda}_{\text{scalar}} w$$

Thus, $\boxed{w \propto S_W^{-1}(\mu_1-\mu_2)}$

The scale doesn't matter because it cancels in the ratio above.