# PCA

## Problem 3.

(a). $\tilde{Y}_1$ preserves 70% variance of $Y_1$. $Cov(X) = \frac{1}{N} X^T X$. $Cov(Y_1) = \lambda^2 \cdot \frac{1}{N} X^T X$

for $Y_1$, $\sum_{i=1}^{k} \lambda_i^{Y_1} = \lambda^2 \sum_{i=1}^{k} \lambda_i^{X} \geq 0.7 \cdot \lambda^2 \sum_{i=1}^{D} \lambda_i^{X} = 0.7 \sum_{i=1}^{D} \lambda_i^{Y_1}$

(b). $\tilde{Y}_2$ preserves 70% variance of $Y_2$

R is an orthonogal matrix, it only rotates the data matrix X

doesn't chang the variance (eigenvalues) of the covariance matrix.

(c). $\tilde{Y}_3$ preserves 70% variance of $Y_3$

$Cov(Y_3) = \frac{1}{N} \cdot (XP)^T \cdot XP = \frac{1}{N} \cdot P^T X^T X P = P^T \Sigma P$

for $Y_3$, $\sum_{i=1}^{k} \lambda_i^{Y_3} = 25 \sum_{i=1}^{k} \lambda_i^{X} \geq 0.7 \times 25 \cdot \sum_{i=1}^{D} \lambda_i^{X} = 0.7 \sum_{i=1}^{D} \lambda_i^{Y_3}$

so $\sum_{i=1}^{k} \lambda_i^{Y_3} \geq 0.7 \cdot \sum_{i=1}^{D} \lambda_i^{Y_3}$

(d). can not tell without additional information. it scales $\overbrace{\text{each}}^{\text{data in}}$ dimension with different number, which has changed the covariance distribution.

(e). $\tilde{Y}_5$ preserves 70% variance of $Y_5$

PCA will center the data, which will remove the shifting amount $\mu$.

(f). $\tilde{Y}_6$ preserves 100% variance of $Y_6$

$\text{rank}(Y_6) \leq 5$. when $k=5$, the projected data has preserved all the dimensions.

## Problem 4.

(a). 1. center the data: $\bar{x} = [\bar{x}_1, \bar{x}_2, \bar{x}_3] = \frac{1}{4}\begin{bmatrix} 4+2+4-2 \\ 3+1-1+1 \\ 2-2+3+2 \end{bmatrix}^T = \frac{1}{4}\begin{bmatrix} 8 \\ 4 \\ 4 \end{bmatrix}^T = [2, 1, 1]$

$$\tilde{X} = X - \bar{x} = \begin{bmatrix} 2 & 2 & 1 \\ 0 & 0 & -3 \\ 2 & -2 & 1 \\ -4 & 0 & 1 \end{bmatrix}$$

2. Compute the covariance matrix:

$$\Sigma_{\tilde{X}} = \frac{1}{N} \cdot \tilde{X}^T \tilde{X} = \frac{1}{4} \cdot \begin{bmatrix} 2 & 0 & 2 & -4 \\ 2 & 0 & -2 & 0 \\ 1 & -3 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 2 & 2 & 1 \\ 0 & 0 & -3 \\ 2 & -2 & 1 \\ -4 & 0 & 1 \end{bmatrix}$$

$$= \frac{1}{4}\begin{bmatrix} 24 & 0 & 0 \\ 0 & 8 & 0 \\ 0 & 0 & 12 \end{bmatrix} = \begin{bmatrix} 6 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

3. Eigenvector Decomposition of $\Sigma_{\hat{x}}$

$$\Sigma_{\hat{x}} = \Gamma^T \Lambda \Gamma = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}^T \begin{bmatrix} 6 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$PC_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad PC_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad PC_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

$$Var(PC_1) = 6 \quad Var(PC_2) = 2 \quad Var(PC_3) = 3$$

(b). the top 2 PC is $PC_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ and $PC_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$

$$\Gamma_{truncated} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}$$

$$Y = \tilde{X} \cdot \Gamma_{truncated} = \begin{bmatrix} 2 & 2 & 1 \\ 0 & 0 & -3 \\ 2 & -2 & 1 \\ -4 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 0 & -3 \\ 2 & 1 \\ -4 & 1 \end{bmatrix}$$

$$\sum_{i=1}^{2} \lambda_i = 6+3 = 9$$

$$\sum_{i=1}^{3} \lambda_i = 6+3+2 = 11$$

$$\text{fraction} = \frac{9}{11} = 81.8\%$$

(c). $x_5$ should not change the mean $\bar{x} = [2, 1, 1]$, or the covariance matrix will change.

so set $x_5 = [2, 1, 1]$

Problem 5. $P = MV \Rightarrow P_{Leslie} = [0, 3, 0, 0, 4] \cdot \begin{bmatrix} 0.58 & 0 \\ 0.58 & 0 \\ 0.58 & 0 \\ 0 & 0.71 \\ 0 & 0.71 \end{bmatrix} = [1.74, 2.84]$

Sci-fi-Concept

Romance-Concept

according to the projected data, Leslie likes the Romance-concept movies more than the Scifi-concept movies, so she will rate higher for Casablanca than the Matrix and Star Wars.

Problem 6. the optimal solution $w^* = (X^T X^{-1}) X^T y$

$X = U \Sigma V^T$

$U \in \mathbb{R}^{N \times D}$

$\Sigma \in \mathbb{R}^{D \times D}$

$V^T \in \mathbb{R}^{D \times D}$

$$\begin{aligned}
w^* &= (U \Sigma V^T)^T (U \Sigma V^T)^{-1} (U \Sigma V^T)^T y \\
&= (V \Sigma U^T \cdot U \Sigma V^T)^{-1} \cdot (V \Sigma U^T) y \\
&= (V \Sigma^2 V^T)^{-1} \cdot V \Sigma U^T \cdot y \\
&= (V^T)^{-1} \Sigma^{-2} \cdot V^{-1} \cdot V \Sigma U^T \cdot y \\
&= V \Sigma^{-2} \cdot \Sigma U^T y \\
&= V \Sigma^{-1} U^T y
\end{aligned}$$

$U^T y$: $O(D * N \times N * 1) = O(D \cdot N)$

$\Sigma^{-1} U^T y$: $O(D*D \cdot D*1)$ $\Sigma$ is diagonal $= O(D)$

$V \Sigma^{-1} U^T y$: $O(D*D \cdot D*1) = O(D^2)$

$N > D$, so the dominant complexity is $O(D \cdot N)$