③ a) $y_1 = XS = X\lambda I = \lambda X$

since there is only a difference in scale between y_1 and X all variance will be preserved : $\frac{1}{N}\sum_{i=0}^{N}(x_i - \bar{x_i})^2 \times 0.7 = \frac{1}{N}\sum_{i=0}^{N}(\tilde{x_i} - \bar{\tilde{x_i}})^2$

$$\frac{1}{N}\sum_{i=0}^{N}(y_i - \bar{y_i})^2 = \frac{1}{N}\cdot\lambda\sum_{i=0}^{N}(x_i - \bar{x_i})$$

→ PCA on the top-5 components will still yield 70% of the variance

considering that D can be interpeted as a rotation of the axis, PCA will still yield 70% of the variance in a rotated space

b) $y_2 = XR$, $R \in {}^{D\times D}$ and $RR^{-1} = I$

- considering that D can be interpreted as a rotation / reflection over an axis, PCA will still yield 70% of the variance in the new space

c) $y_3 = XP$ with $P = diag(5, -5, 5, -5, ---, -5)$

- considering 5 is just a scaling factor we can ignore it (a))
- considering $P^T = P$ and $\frac{1}{25}PP = I$ we encounter the same situation as b)

→ 70% of the variance is kept.

d) $y_4 = XQ$ with $Q = diag(1, 2, ---, D)$

- $QQ \neq I$ such that this is not a reflection / rotation
- $\Sigma_{y_4} = \frac{1}{N}Q^TX^TXQ$ in which the result isn't clean and we can't tell the change in variance without additional information

→ every axis is scaled differently.

e) $y_5 = X + 1_N\mu^T$ where $\mu \in \mathbb{R}^D$

$$1_N\mu^T = N\left\{\begin{bmatrix} \overset{\mu}{\text{———}} \\ \text{———} \\ \vdots \\ \text{———} \end{bmatrix}\right.$$

we shift every data point by μ but the first step of PCA is removing the mean, thus nullifying the shift operation and yielding the same 70%

f) $Y_4 = XA$, rank$(A) = 5$

⇒ since the rank of a matrix multiplication is the lowest of the two, rank$(y_4) \leq 5$ (in case rank$(X) < 5 \Rightarrow$ rank$(XA) < 5$)

⇒ there will be no more than 5 eigenvalues in the PCA decomposition, so keeping the top 5 will yield 100 % of the variance.

④ a) $X = \begin{pmatrix} 4 & 3 & 2 \\ 2 & 1 & -2 \\ 4 & -1 & 2 \\ -2 & 1 & 2 \end{pmatrix}$, $\bar{x} = \frac{1}{N} X^T \cdot 1_N = \frac{1}{N}(8\ 4\ 4)^T = (2\ 1\ 1)^T$

$$\tilde{X} = \begin{pmatrix} 2 & 2 & 1 \\ 0 & 0 & -3 \\ 2 & -2 & 1 \\ -4 & 0 & 1 \end{pmatrix} \multimap \Sigma_{\tilde{X}} = \frac{1}{N}\tilde{X}^T\tilde{X} = \frac{1}{4}\begin{pmatrix} 24 & 0 & 0 \\ 0 & 8 & 0 \\ 0 & 0 & 12 \end{pmatrix} = \begin{pmatrix} 6 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

- $\det(\Sigma_{\tilde{X}} - \lambda I) = 0 \Leftrightarrow x = 6 \vee \lambda = 2 \vee \lambda = 3$

- $(\Sigma_{\tilde{X}} - 6I)\begin{pmatrix} v \\ y \\ z \end{pmatrix} = 0 \Leftrightarrow \begin{cases} \overline{} \\ -4y = 0 \\ -3z = 0 \end{cases}$ $\multimap$ eigenvector: $(1\ 0\ 0)$

- $(\Sigma_{\tilde{X}} - 2I)\begin{pmatrix} v \\ y \\ z \end{pmatrix} = 0 \Leftrightarrow \begin{cases} 4v = 0 \\ \overline{} \\ z = 0 \end{cases}$ $\multimap$ eigenvector: $(0\ 1\ 0)$

- $(\Sigma_{\tilde{X}} - 3I)\begin{pmatrix} v \\ y \\ z \end{pmatrix} = 0 \Leftrightarrow \begin{cases} 3x = 0 \\ -y = 0 \end{cases}$ $\multimap$ eigenvector: $(0\ 0\ 1)$

- var along $(1\ 0\ 0)$: $\frac{\Sigma(x_i - \bar{x}_i)^2}{N} = \frac{1+4+16}{4} = \textcircled{6}$
- " " $(0\ 1\ 0)$: 2
- " " $(0\ 0\ 1)$ ③

b) $y = \tilde{X}\,\Gamma_{\text{truncated}} = \begin{pmatrix} 2 & 2 & 1 \\ 0 & 0 & -3 \\ 2 & -2 & 1 \\ -4 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 0 & -3 \\ 2 & 1 \\ -4 & 1 \end{pmatrix}$

↳ var = 6

→ var = 3

⇒ fraction of the variance conserved:

$$\frac{6+3}{6+2+3} = 9/11 \simeq 0{,}82$$

c) consider a new entry which preserves the PCA analysis.

- It would have to conserve the mean and result in a similar Σ

- consider $x_5 = \bar{x} = (2\ 1\ 1)$

  obviously the mean stays the same

- $\Sigma_{\bar{x}}' = \frac{1}{5}\begin{pmatrix} 24 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 12 \end{pmatrix}$ → similar to $\Sigma_{\bar{x}}$ (change of scale)

  since the eigenvectors will stay equal in a different scale we will encounter

  precisely the same principal components when adding $x_5 = \bar{x}$ to the dataset

⑤ $P = MV$

$\Rightarrow$ Leslie: $\begin{pmatrix} 0 & 3 & 0 & 0 & 4 \end{pmatrix}\begin{pmatrix} 0{,}58 & 0 \\ 0{,}58 & 0 \\ 0{,}58 & 0 \\ 0 & 0{,}71 \\ 0 & 0{,}71 \end{pmatrix} = \begin{pmatrix} 1{,}74 & 2{,}84 \end{pmatrix}$

→ it seems like Leslie prefers the second concept present in our data so we should recommend movies similar to

the Titanic, Casablanca seems like a reasonable choice.

⑥ $\cdot\ Xw = y \rightarrow U\Sigma V^T w = y$

$$\Rightarrow (U\Sigma V^T)^{-1}(U\Sigma V^T)w = (U\Sigma V^T)^{-1}y$$

pseudo inverse ↖

$$\Rightarrow V\Sigma^{-1}U^T U\Sigma V^T w = (U\Sigma V^T)^{-1}y$$

$$\Rightarrow w^* = V\Sigma^{-1}U^T y$$

$$w^* = V\Sigma^{-1}U^T y$$

$\Rightarrow V \in M^{d\times d}$

$\Rightarrow \Sigma^{-1} \in M^{d\times d}$

$\Rightarrow U^T \in M^{d\times m}$

$\Rightarrow y \in K^N$

$X = U\Sigma V^+$

$m\times d \quad d\times d \quad d\times d$

(X is full rank → number of singular values is d, assuming $d < m$)

→ computing w^* starting from the right:

- $U^T y \rightarrow O(dm)$
- $\Sigma^{-1}(U^T y) \rightarrow O(d) + O(dm)$
- $V(\underbrace{\Sigma^{-1}U^T y}_{\in \mathbb{R}^d}) \rightarrow O(d^2)$

$\Bigg\}\ O(dm) + O(d) + O(dm) + O(d^2) = O(md)$ //

(assuming $d < m$)