Week 9: Revision

Q1. A dataset has 100 observations. Each observation is in $\mathbb{R}^7$. What is the dimension of the scatter matrix of this dataset?

$7 \times 100 \qquad 100 \times 7$

$\therefore$ The dimension of the scatter matrix is 7×7

Q2. In this question, your answer 4 d.p.

let $x = \begin{bmatrix} 3 \\ 3 \\ -4 \end{bmatrix}$, $u_1 = \begin{bmatrix} 2 \\ 3 \\ -1 \end{bmatrix}$ and $u_2 = \begin{bmatrix} 4 \\ -3 \\ -1 \end{bmatrix}$

Then $\langle x, u_1 \rangle$? and $\langle x, u_2 \rangle$?

- Let the projection of x on the subspace spanned by u_1, u_2 be $\begin{bmatrix} a \\ b \\ c \end{bmatrix}$. Then $a = ?$, $b = ?$ and $c = ?$

$\langle x, u_1 \rangle = 3.2 + 3.3 + (-4).(-1) = 6 + 9 + 4 = 19$

$\langle x, u_2 \rangle = 3.4 + 3.(-3) + (-4).(-1) = 12 - 9 + 4 = 7$

First, check that u_1 and u_2 are orthogonal.

$\langle u_1, u_2 \rangle = 2.4 + 3.(-3) + (-1).(-1) = 8 - 9 + 1 = 0$

- Then, compute $\dfrac{\langle x, u_1 \rangle}{\langle u_1, u_1 \rangle}$ and $\dfrac{\langle x, u_2 \rangle}{\langle u_2, u_2 \rangle}$

$\langle u_1, u_1 \rangle = 2.2 + 3.3 + (-1).(-1) = 4 + 9 + 1 = 14$

$\langle u_2, u_2 \rangle = 4.4 + (-3).(-3) + (-1).(-1) = 16 + 9 + 1 = 26$

$$\text{Projection} = \frac{\langle x, u_1 \rangle}{\langle u_1, u_1 \rangle} . u_1 + \frac{\langle x, u_2 \rangle}{\langle u_2, u_2 \rangle} . u_2$$

$$= \frac{19}{14} . \begin{bmatrix} 2 \\ 3 \\ -1 \end{bmatrix} + \frac{7}{26} . \begin{bmatrix} 4 \\ -3 \\ -1 \end{bmatrix} = \begin{bmatrix} \frac{19}{7} \\ \frac{57}{14} \\ -19/14 \end{bmatrix} + \begin{bmatrix} 14/13 \\ -21/26 \\ -7/26 \end{bmatrix}$$

$$= \begin{bmatrix} 3.7912 \\ 3.2637 \\ -1.6264 \end{bmatrix}$$

Q3. A dataset was pre-processed for PCA. After pre-processing, the four observations in the dataset are $\begin{bmatrix}1\\-4\end{bmatrix}, \begin{bmatrix}2\\-2\end{bmatrix}, \begin{bmatrix}-3\\5\end{bmatrix}$ & $\begin{bmatrix}0\\1\end{bmatrix}$

Let $\begin{bmatrix}a & b\\ b & c\end{bmatrix}$ be the scatter matrix of this dataset.

Then $a=?$, $b=?$ and $c=?$

Let $x_1=\begin{bmatrix}1\\-4\end{bmatrix}$ $x_2=\begin{bmatrix}2\\-2\end{bmatrix}$ $x_3=\begin{bmatrix}-3\\5\end{bmatrix}$ $x_4=\begin{bmatrix}0\\1\end{bmatrix}$

$$x_1(x_1)'=\begin{bmatrix}1\\-4\end{bmatrix}\begin{bmatrix}1 & -4\end{bmatrix}=\begin{bmatrix}1\times1 & 1\times-4\\ -4\times1 & -4\times-4\end{bmatrix}=\begin{bmatrix}1 & -4\\ -4 & 16\end{bmatrix}.$$

$2\times\boxed{1}$ $\boxed{1}\times 2$

$$x_2(x_2)'=\begin{bmatrix}2\\-2\end{bmatrix}\begin{bmatrix}2 & -2\end{bmatrix}=\begin{bmatrix}2\times2 & 2\times-2\\ -2\times2 & -2\times-2\end{bmatrix}=\begin{bmatrix}4 & -4\\ -4 & 4\end{bmatrix}.$$

$$x_3(x_3)'=\begin{bmatrix}-3\\5\end{bmatrix}\begin{bmatrix}-3 & 5\end{bmatrix}=\begin{bmatrix}-3\times-3 & -3\times5\\ 5\times-3 & 5\times5\end{bmatrix}=\begin{bmatrix}9 & -15\\ -15 & 25\end{bmatrix}$$

$$x_4(x_4)'=\begin{bmatrix}0\\1\end{bmatrix}\begin{bmatrix}0 & 1\end{bmatrix}=\begin{bmatrix}0\times0 & 0\times1\\ 1\times0 & 1\times1\end{bmatrix}=\begin{bmatrix}0 & 0\\ 0 & 1\end{bmatrix}$$

$$\sum_{i=1}^{4} x_i(x_i)'=\begin{bmatrix}1 & -4\\ -4 & 16\end{bmatrix}+\begin{bmatrix}4 & -4\\ -4 & 4\end{bmatrix}+\begin{bmatrix}9 & -15\\ -15 & 25\end{bmatrix}+\begin{bmatrix}0 & 0\\ 0 & 1\end{bmatrix}$$

$$=\begin{bmatrix}1+4+9+0 & -4-4-15+0\\ -4-4-15+0 & 16+4+25+1\end{bmatrix}=\begin{bmatrix}14 & -23\\ -23 & 46\end{bmatrix}$$

Q4. In this question, 3 d.p.
It is known that the matrix $\begin{bmatrix} 1 & 2 & 6 \\ 2 & 6 & 1 \\ 6 & 1 & 4 \end{bmatrix}$
has an eigenvector with corresponding eigenvalue equals to 5.
If the eigenvector is $\begin{bmatrix} 1 \\ a \\ b \end{bmatrix}$, then a = ? and b = ?

$$A \cdot u = \lambda \cdot u$$

where,

A is a dxd square matrix
u $u \in \mathbb{R}^d$ is an eigenvector of A
λ is eigenvalue of the eigenvector.

let $A = \begin{bmatrix} 1 & 2 & 6 \\ 2 & 6 & 1 \\ 6 & 1 & 4 \end{bmatrix}$ $u = \begin{bmatrix} 1 \\ a \\ b \end{bmatrix}$ $\lambda = 5$

$$\underset{3\times3}{\begin{bmatrix} 1 & 2 & 6 \\ 2 & 6 & 1 \\ 6 & 1 & 4 \end{bmatrix}} \cdot \underset{3\times1}{\begin{bmatrix} 1 \\ a \\ b \end{bmatrix}} = 5 \cdot \begin{bmatrix} 1 \\ a \\ b \end{bmatrix} = \begin{bmatrix} 5 \\ 5a \\ 5b \end{bmatrix}$$

$$\begin{bmatrix} 1\times1 + 2\times a + 6\times b \\ 2\times1 + 6\times a + 1\times b \\ 6\times1 + 1\times a + 4\times b \end{bmatrix} = \begin{bmatrix} 1+2a+6b \\ 2+6a+b \\ 6+a+4b \end{bmatrix} = \begin{bmatrix} 5 \\ 5a \\ 5b \end{bmatrix}$$

$1 + 2a + 6b = 5$
$2a + 6b = 4$

$2 + 6a + b = 5a$
$a + b = -2$
$a = -2 - b$
$a = -2 - 2$
$a = -4$

$6 + a + 4b = 5b$
$a - b = -6$
$-2 - b - b = -6$
$-2b = -4$
$b = \frac{-4}{-2} = 2.$

Week 9 : Quiz

Q1. The dimension of a matrix A is 7x15. What is the dimension of its transpose A'?

Solution: 15 x 7

Q2. The key motivation behind PCA is to seek a vector u that maximizes the variance of the projections of the observations onto the subspace spanned by u. As it turns out, u is the eigenvector of the scatter matrix with the largest eigenvalue.

Q3. 4 d.p. Let $x = \begin{bmatrix} -1 \\ 0 \\ 5 \end{bmatrix}$, $u_1 = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$ and $u2 = \begin{bmatrix} 4 \\ -3 \\ -2 \end{bmatrix}$

Then $\langle x, u1 \rangle = ?$ and $\langle x, u2 \rangle = ?$

Let the projection of x on the subspace spanned by $u1, u2$ be $\begin{bmatrix} a \\ b \\ c \end{bmatrix}$. Then $a = ?$, $b = ?$ and $c = ?$

$\langle x, u1 \rangle = (-1) \cdot 1 + 0 \cdot 2 + 5 \cdot (-1) = -1 + 0 + (-5) = -6$

$\langle x, u2 \rangle = (-1) \cdot 4 + 0 \cdot (-3) + 5 \cdot (-2) = -4 + 0 - 10 = -14$

First, check that u_1 and u_2 are orthogonal

$\langle u_1, u_2 \rangle = 1 \cdot 4 + 2 \cdot (-3) + (-1) \cdot (-2) = 4 - 6 + 2 = 0$

Then, compute $\dfrac{\langle x, u_1 \rangle}{\langle u_1, u_1 \rangle}$ and $\dfrac{\langle x, u_2 \rangle}{\langle u_2, u_2 \rangle}$

$\langle u_1, u_1 \rangle = (1)^2 + 2^2 + (-1)^2 = 1 + 4 + 1 = 6.$

$\langle u_2, u_2 \rangle = 4^2 + (-3)^2 + (-2)^2 = 16 + 9 + 4 = 29$

$$\text{Projection} = \frac{-6}{6} \cdot \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} + \frac{-14}{29} \cdot \begin{bmatrix} 4 \\ -3 \\ -2 \end{bmatrix} = \begin{bmatrix} -1 \\ -2 \\ 1 \end{bmatrix} + \begin{bmatrix} -56/29 \\ 42/29 \\ 28/29 \end{bmatrix}$$

$$= \begin{bmatrix} -85/29 \\ -16/29 \\ 57/29 \end{bmatrix} = \begin{bmatrix} -2.9310 \\ -0.5517 \\ 1.9655 \end{bmatrix}$$

Week 9: Quiz

Q4. A dataset was pre-processed for PCA. After pre-processing, the three observations in the dataset are

$$\begin{bmatrix} 2 \\ -2 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ -1 \end{bmatrix} \text{ and } \begin{bmatrix} -3 \\ -1 \\ 2 \end{bmatrix}. \text{ Let } \begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix} \text{ be the scatter matrix}$$

of this dataset. Then $a = ?$, $b = ?$ and $e = ?$

$$\text{Let } x_1 = \begin{bmatrix} 2 \\ -2 \\ -1 \end{bmatrix}, x_2 = \begin{bmatrix} 1 \\ 3 \\ -1 \end{bmatrix} \text{ and } x_3 = \begin{bmatrix} -3 \\ -1 \\ 2 \end{bmatrix}$$

$$x_1(x_1)' = \underset{3\times 1}{\begin{bmatrix} 2 \\ -2 \\ -1 \end{bmatrix}} \underset{1\times 3}{\begin{bmatrix} 2 & -2 & -1 \end{bmatrix}} = \begin{bmatrix} 2\times2 & 2\times-2 & 2\times-1 \\ -2\times2 & -2\times-2 & -2\times-1 \\ -1\times2 & -1\times-2 & -1\times-1 \end{bmatrix} = \underset{3\times3}{\begin{bmatrix} 4 & -4 & -2 \\ -4 & 4 & 2 \\ -2 & 2 & 1 \end{bmatrix}}$$

$$x_2(x_2)' = \begin{bmatrix} 1 \\ 3 \\ -1 \end{bmatrix} \begin{bmatrix} 1 & 3 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 3 & -1 \\ 3 & 9 & -3 \\ -1 & -3 & 1 \end{bmatrix}$$

$$x_3(x_3)' = \begin{bmatrix} -3 \\ -1 \\ 2 \end{bmatrix} \begin{bmatrix} -3 & -1 & 2 \end{bmatrix} = \begin{bmatrix} 9 & 3 & -6 \\ 3 & -1 & -2 \\ -6 & -2 & 4 \end{bmatrix}$$

$$\sum_{i=1}^{3} x_i(x_i)' = \begin{bmatrix} 14 & 2 & -9 \\ 2 & 12 & -3 \\ -9 & -3 & 6 \end{bmatrix}$$

Q5. For this question, you may use R or Python, accurat to 4 d.p.

$$\text{What is the sum of all eigenvalues of } \begin{bmatrix} 13.3 & -6.4 & 1.8 & -17.6 \\ -6.4 & -7.7 & 1.8 & -10.1 \\ 1.8 & 1.8 & -4.4 & 27.9 \\ -17.6 & -10.1 & 27.9 & 5.2 \end{bmatrix}$$

In R:

```
q <- matrix(c(put the dataset here), ncol=4, byrow=TRUE)

eigen_of_q <- eigen(q)
eigen_of_q$values
sum(eigen_of_q$values).
```