Reference Solution of COMP7180 Quiz1

Question 1 (22 Marks)

(a) (10 marks) Consider a linear equation system $\mathbf{Ax} = \mathbf{b}$, where $\mathbf{A} = \begin{bmatrix} 1 & -2 \\ 3 & a \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 1 \\ b \end{bmatrix}$. Find all the values of a and b so that this system has **NO** solution. Justify your answers.

(b) (12 marks) Which of the following matrices is(are) guaranteed to equal $(\mathbf{A} - \mathbf{B})^2$?

$\mathbf{A}^2 - \mathbf{B}^2$, $(\mathbf{B} - \mathbf{A})^2$, $\mathbf{A}^2 - 2\mathbf{AB} + \mathbf{B}^2$, $\mathbf{A}^2 - \mathbf{AB} - \mathbf{BA} + \mathbf{B}^2$

Justify your answers.

Reference Solution:

(a) We write down the linear equation system as

$$\begin{cases} 1\,x_1 - 2\,x_2 = 1 \quad --(1) \\ 3\,x_1 + a\,x_2 = b \quad --(2) \end{cases}$$

We perform the row operation : $equation\ (2) - 3 \times\ (1)$, then we have:

$$\begin{cases} 1\,x_1 - 2\,x_2 \qquad\quad = 1 \qquad --(1) \\ 0\,x_1 + (a+6)x_2 = b - 3 \quad --(2) \end{cases}$$

The linear system has no solution, when the equation can not hold. This equation will be contradictory if:

$$a + 6 = 0 \text{ and } b - 3 \neq 0$$

As a result, the linear system has no solution when $a = -6$ and $b \neq 3$.

(b) We can guarantee $(\mathbf{B} - \mathbf{A})^2$ and $\mathbf{A}^2 - \mathbf{AB} - \mathbf{BA} + \mathbf{B}^2$

For formula $(\mathbf{A} - \mathbf{B})^2$, we have

$$(\mathbf{A} - \mathbf{B})^2 = (\mathbf{A} - \mathbf{B})(\mathbf{A} - \mathbf{B}) = \mathbf{A}^2 - \mathbf{AB} - \mathbf{BA} + \mathbf{B}^2$$

As the exchange law is not valid in matrix multiplication, $\mathbf{AB} + \mathbf{BA} \neq \mathbf{2\,AB}$.

Also, we have $(\mathbf{B} - \mathbf{A})^2 = \mathbf{B}^2 - \mathbf{BA} - \mathbf{AB} + \mathbf{A}^2 = \mathbf{A}^2 - \mathbf{AB} - \mathbf{BA} + \mathbf{B}^2 = (\mathbf{A} - \mathbf{B})^2$

Question 2 (30 Marks)

(a) (10 marks) Given three n-dimensional vectors: $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$. If ($3\mathbf{x}$) and ($2\mathbf{x}+\mathbf{y}$) and ($\mathbf{x}+\mathbf{y}+\mathbf{z}$) are linearly independent, do we have the conclusion that $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ are also linearly independent? Justify your answer.

(b) (20 marks) Given three vectors: $\boldsymbol{x} = (1,0,0)^T, \boldsymbol{y} = (y_1, y_2, y_3)^T, \boldsymbol{z} = (z_1, z_2, z_3)^T$. Suppose both $\boldsymbol{y}$ and $\boldsymbol{z}$ are unit vectors and all the elements in these two vectors are real numbers. Moreover, we know that $\boldsymbol{x}$ and $\boldsymbol{y}$ are orthogonal. What are the minimum and maximum values of $(\boldsymbol{x}^T\boldsymbol{z})^2 + (\boldsymbol{y}^T\boldsymbol{z})^2$? Justify your answer.

Reference Solution:

(a) Because ($3\mathbf{x}$) and ($2\mathbf{x}+\mathbf{y}$) and ($\mathbf{x}+\mathbf{y}+\mathbf{z}$) are linearly independent, for the equation:

$$\mathrm{c}_1\,(3\mathbf{x}) + \mathrm{c}_2\,(2\mathbf{x}+\mathbf{y}) + \mathrm{c}_3\,(\mathbf{x}+\mathbf{y}+\mathbf{z}) = 0$$

the only solution is $c_1 = c_2 = c_3 = 0$, which also implies that the only solution for

$$(3\mathrm{c}_1 + 2\mathrm{c}_2 + \mathrm{c}_3)\mathbf{x} + (\mathrm{c}_2 + c_3)\mathbf{y} + \mathrm{c}_3\,\mathbf{z} = 0$$

is $c_1 = c_2 = c_3 = 0$, where $(3\mathrm{c}_1 + 2\mathrm{c}_2 + \mathrm{c}_3) = (\mathrm{c}_2 + c_3) = \mathrm{c}_3 = 0$.

Therefore, vectors $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ are also linearly independent.

(b) Consider there exist a unit vector $\boldsymbol{w} = (w_1, w_2, w_3)^T$ are orthogonal to both vectors $\boldsymbol{x}$ and $\boldsymbol{y}$. Therefore, we have an orthonormal basis $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w}$ in space R^3. We can represent vector $\boldsymbol{z}$ by the linear combination of vectors $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w}$ as $\boldsymbol{z} = a\boldsymbol{x} + b\boldsymbol{y} + c\boldsymbol{w}$.

Then $\boldsymbol{x}^\top\boldsymbol{z} = \boldsymbol{x}^\top(a\boldsymbol{x} + b\boldsymbol{y} + c\boldsymbol{w})$, $\boldsymbol{y}^\top\boldsymbol{z} = \boldsymbol{y}^\top(a\boldsymbol{x} + b\boldsymbol{y} + c\boldsymbol{w})$.

As $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w}$ is an orthonormal basis, we have $\boldsymbol{x}^\top\boldsymbol{y} = \boldsymbol{x}^\top\boldsymbol{w} = \boldsymbol{y}^\top\boldsymbol{x} = \boldsymbol{y}^\top\boldsymbol{w} = 0$.

As $\boldsymbol{x}$ and $\boldsymbol{y}$ are unit vectors, we have $\boldsymbol{x}^\top\boldsymbol{x} = \boldsymbol{y}^\top\boldsymbol{y} = 1$.

Thus,

$$\boldsymbol{x}^\top\boldsymbol{z} = \boldsymbol{x}^\top(a\boldsymbol{x} + b\boldsymbol{y} + c\boldsymbol{w}) = a\boldsymbol{x}^\top\boldsymbol{x} + b\boldsymbol{x}^\top\boldsymbol{y} + c\boldsymbol{x}^\top\boldsymbol{w} = a\boldsymbol{x}^\top\boldsymbol{x} = a$$
$$\boldsymbol{y}^\top\boldsymbol{z} = \boldsymbol{y}^\top(a\boldsymbol{x} + b\boldsymbol{y} + c\boldsymbol{w}) = a\boldsymbol{y}^\top\boldsymbol{x} + b\boldsymbol{y}^\top\boldsymbol{y} + c\boldsymbol{y}^\top\boldsymbol{w} = b\boldsymbol{y}^\top\boldsymbol{y} = b$$
$$(\boldsymbol{x}^\top\boldsymbol{z})^2 + (\boldsymbol{y}^\top\boldsymbol{z})^2 = a^2 + b^2$$

Because $\boldsymbol{z}$ is a unit vector, $a^2 + b^2 + c^2 = 1$.

We have $a^2 + b^2 = 0$, when $c^2 = 1$ (vector $\boldsymbol{z}$ is on the same as vector $\boldsymbol{w}$, orthogonal to both vectors $\boldsymbol{x}, \boldsymbol{y}$).

We have $a^2 + b^2 = 1$, when $c^2 = 0$ (vector $\boldsymbol{z}$ is orthogonal to vector $\boldsymbol{w}$, on the plane that is spanned by to vectors $\boldsymbol{x}, \boldsymbol{y}$).

Therefore, $0 \le (\boldsymbol{x}^T\boldsymbol{z})^2 + (\boldsymbol{y}^T\boldsymbol{z})^2 \le 1$.

The minimum can be obtained when vector $\boldsymbol{z}$ are orthogonal to both vectors $\boldsymbol{x}, \boldsymbol{y}$.

The maximum can be obtained when vector $\boldsymbol{z}$ are on the plane that is spanned by to vectors $\boldsymbol{x}, \boldsymbol{y}$.

Question 3 (24 Marks)

(a) (10 marks) Suppose that $\mathbf{P}_1$ is the projection matrix onto the line through (1, 2) and $\mathbf{P}_2$ is the projection matrix onto the line through (2, 1). Compute $\mathbf{P} = \mathbf{P}_2\mathbf{P}_1$. Show your calculation details.

(b) (14 marks) Given the following matrix:

$$\mathrm{A} = \begin{bmatrix} 1 & 1 & 3 \\ 1 & 2 & a \\ 2 & 2 & 4 \end{bmatrix}.$$

There is another matrix **B** that satisfies the following equality:

$\mathbf{AB} = \mathbf{A} - \mathbf{B} + \mathbf{I}$, where **I** denotes the 3×3 identity matrix.

Moreover, we know that $\mathbf{B} \neq \mathbf{I}$.

Find the value of the variable a in the matrix **A**. Show your calculation details.

Reference Solution:

(a) Consider $a = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, b = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$, we have $\mathbf{P_1} = \frac{a\,a^\top}{a^\top a} = \frac{1}{5}\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$, $\mathbf{P_2} = \frac{b\,b^\top}{b^\top b} = \frac{1}{5}\begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix}$.

Then, we have $\mathbf{P} = \mathbf{P}_2\mathbf{P}_1 = \frac{\mathbf{1}}{\mathbf{25}}\begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} = \frac{\mathbf{1}}{\mathbf{25}}\begin{bmatrix} 8 & 16 \\ 4 & 8 \end{bmatrix}$

(b)From $\mathbf{AB} = \mathbf{A} - \mathbf{B} + \mathbf{I}$, we have

$$\mathbf{AB} + \mathbf{B} = (\mathbf{A} + \mathbf{I})\mathbf{B} = \mathbf{A} + \mathbf{I}$$

If $(\mathbf{A} + \mathbf{I})$ is inversible , then $\boldsymbol{B} = (\mathbf{A} + \mathbf{I})^{-\mathbf{1}}(\mathbf{A} + \mathbf{I})\mathbf{B} = (\mathbf{A} + \mathbf{I})^{-\mathbf{1}}(\mathbf{A} + \mathbf{I}) = \mathbf{I}.$

But $\mathbf{B} \neq \mathbf{I}$. So, $(\mathbf{A} + \mathbf{I})$ is not inversible. Therefore, $(\mathbf{A} + \mathbf{I})$ can not be a full rank matrix.

$$\mathbf{A} + \mathbf{I} = \begin{bmatrix} \mathbf{2} & \mathbf{1} & \mathbf{3} \\ \mathbf{1} & \mathbf{3} & \boldsymbol{a} \\ \mathbf{2} & \mathbf{2} & \mathbf{5} \end{bmatrix}$$

We only need to find the linear combination of $\begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$ to produce $\begin{bmatrix} 3 \\ a \\ 5 \end{bmatrix}$.

We solve the equation system $c_1\begin{bmatrix} 2 \\ 2 \end{bmatrix} + c_2\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 5 \end{bmatrix}$. And we have the result with $c_1 = \frac{1}{2}, c_2 = 2$.

Therefore, $a = 1c_1 + 3c_2 = \frac{1}{2} + 6 = \frac{13}{2}$.

Question 4 (24 Marks)

(a) (12 marks) Principal Component Analysis (PCA) is a well-known dimensionality reduction algorithm. It aims to find a low-dimensional representation of the high-dimensional data while keeping the variance of the given dataset. However, if the original dimension (denoted by D) is extremely high, PCA might be computationally expensive or even infeasible as the size of the covariance matrix ($D \times D$) is extremely large. For instance, in a high-resolution video retrieval application, the original dimension of the video data could be hundreds of billions, which is much larger than the number of videos to be analyzed (denoted by n). Suggest a way to reduce the computational cost of PCA under this circumstance.

(b) (12 marks) Given the $D \times n$ data matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n]$, where D is the dimension of data points $\mathbf{x}_1$, $\mathbf{x}_2$, ..., $\mathbf{x}_n$, and n is the number of the data points. Assume that the dataset is centered, i.e., $\bar{\mathbf{x}} = (\mathbf{x}_1 + \mathbf{x}_2 + ... + \mathbf{x}_n) / n = 0$, and the first and second principal component of $\mathbf{X}$ learned by PCA are denoted by $\mathbf{a}_1$ and $\mathbf{a}_2$, derive the **third principal component** of **X,** i.e., $\mathbf{a}_3$, learned by PCA. Show your calculation details.

Reference Solution:

(a) As the dimension D is much larger than number of videos n, the covariance matrix $\boldsymbol{S} = \frac{1}{n}\mathbf{X_c}\mathbf{X_c}^\top$ is a $D \times D$ matrix. It would be difficult to perform eigen-decomposition on matrix $\boldsymbol{S}$.

The main reason of computationally expensive is the large size of matrix $\boldsymbol{S}$. We may try to perform eigen-decomposition to another matrix, which has smaller size than matrix $\boldsymbol{S}$, but with the same eigenvalue.

Here, we propose another matrix $\boldsymbol{S}' = \frac{1}{n}\mathbf{X_c}^\top\mathbf{X_c}$, which is a $n \times n$ matrix. We first prove that the matrices $\boldsymbol{S}$ and $\boldsymbol{S}'$ has the same non-zeros eigenvalue:

- Assume $\lambda, \boldsymbol{v}$ are the eigenvalue and eigenvector of matrix S' , then we have $\frac{1}{n}\mathbf{X_c}^\top\mathbf{X_c}\,\boldsymbol{v} = \lambda\boldsymbol{v}$.
- Assume $\boldsymbol{u} = \mathbf{X_c}\,\boldsymbol{v}$, then we have $\frac{1}{n}\mathbf{X_c}\mathbf{X_c}^\top\mathbf{X_c}\,\boldsymbol{v} = \frac{1}{n}\mathbf{X_c}\mathbf{X_c}^\top\boldsymbol{u} = \boldsymbol{S}\boldsymbol{u} = \lambda\mathbf{X_c}\boldsymbol{v} = \lambda\boldsymbol{u}$.
- As a result, the matrices $\boldsymbol{S}$ and $\boldsymbol{S}'$ has the same non-zeros eigenvalue, and the eigenvector of matrix S is $\boldsymbol{u} = \mathbf{X_c}\,\boldsymbol{v}$, where $\boldsymbol{v}$ is eigenvector of matrix $\boldsymbol{S}'$.

Therefore, we can perform eigen-decomposition to another matrix $\boldsymbol{S}' = \frac{1}{n}\mathbf{X_c}^\top\mathbf{X_c}$, which is a $n \times n$ matrix. And take first k leading eigenvectors $\{\mathbf{X_c}\,\boldsymbol{v_1}, \mathbf{X_c}\,\boldsymbol{v_2}, \dots, \mathbf{X_c}\,\boldsymbol{v_k}\}$ with k largest eigenvalue $\{\boldsymbol{\lambda_1}, \boldsymbol{\lambda_2}, \dots, \boldsymbol{\lambda_k}\}$.

(b) Consider the covariance matrix as $\mathbf{S} = \frac{1}{n}\mathbf{X}\mathbf{X}^T$.

We derive the third principal component of **X**, i.e., $\boldsymbol{a_3}$ by solving the following problem:

$$\max_{\boldsymbol{a}_3} \boldsymbol{a}_3^\top \mathbf{S}\mathbf{a}_3$$

$$s.t.\ \boldsymbol{a}_3^\top\mathbf{a}_3 = 1, \boldsymbol{a}_3^\top\mathbf{a}_1 = 0, \boldsymbol{a}_3^\top\mathbf{a}_2 = 0$$

Then, we maximize the Lagrangian function of the above problem:

$$\max_{\boldsymbol{a}_3} L(\boldsymbol{a}_3) = \max_{\boldsymbol{a}_3} \boldsymbol{a}_3^\top\mathbf{S}\mathbf{a}_3 - \lambda(\boldsymbol{a}_3^\top\mathbf{a}_3 - 1) - \phi_1\boldsymbol{a}_3^\top\mathbf{a}_1 - \phi_2\boldsymbol{a}_3^\top\mathbf{a}_2$$

Then, we take derivatives of Lagrangian functions:

$$\frac{\partial}{\partial\, \boldsymbol{a}_3} \boldsymbol{a}_3^\top \mathbf{S}\mathbf{a}_3 - \lambda(\boldsymbol{a}_3^\top \mathbf{a}_3 - 1) - \phi_1 \boldsymbol{a}_3^\top \mathbf{a}_1 - \phi_2 \boldsymbol{a}_3^\top \mathbf{a}_2 = 2\mathbf{S}\mathbf{a}_3 - 2\lambda \boldsymbol{a}_3 - \phi_1 \boldsymbol{a}_1 - \phi_2 \boldsymbol{a}_2$$

We set the derivative to 0 to find the maximum value:

$$2\mathbf{S}\mathbf{a}_3 - 2\lambda \boldsymbol{a}_3 - \phi_1 \boldsymbol{a}_1 - \phi_2 \boldsymbol{a}_2 = 0$$

As $\boldsymbol{a}_1^\top \boldsymbol{a}_2 = \boldsymbol{a}_1^\top \boldsymbol{a}_3 = 0$, if we left multiply $\boldsymbol{a}_1^\top$ to above equation, we have

$$2\boldsymbol{a}_1^\top \mathbf{S}\mathbf{a}_3 - 2\lambda \boldsymbol{a}_1^\top \boldsymbol{a}_3 - \phi_1 \boldsymbol{a}_1^\top \boldsymbol{a}_1 - \phi_2 \boldsymbol{a}_1^\top \boldsymbol{a}_2 = 0 - 0 - \phi_1 - 0 = 0$$

Therefore, $\phi_1 = 0$.

For ϕ_2, As $\boldsymbol{a}_2^\top \boldsymbol{a}_1 = \boldsymbol{a}_2^\top \boldsymbol{a}_3 = 0$, if we left multiply $\boldsymbol{a}_2^\top$ to above equation, we have

$$2\boldsymbol{a}_2^\top \mathbf{S}\mathbf{a}_3 - 2\lambda \boldsymbol{a}_2^\top \boldsymbol{a}_3 - \phi_1 \boldsymbol{a}_2^\top \boldsymbol{a}_1 - \phi_2 \boldsymbol{a}_2^\top \boldsymbol{a}_2 = 0 - 0 - 0 - \phi_2 = 0$$

Therefore, $\phi_2 = 0$.

Then, we have:

$$2\mathbf{S}\mathbf{a}_3 - 2\lambda \boldsymbol{a}_3 = 0$$

$$\mathbf{S}\mathbf{a}_3 = \lambda \boldsymbol{a}_3$$

As $\boldsymbol{a}_1, \boldsymbol{a}_2$ are eigenvectors corresponding to the first and second largest eigenvalues. Thus, $\boldsymbol{a}_3$ that can maximize $\boldsymbol{a}_3^\top \mathbf{S}\mathbf{a}_3$ and satisfy $\boldsymbol{a}_3^\top \mathbf{a}_1 = 0, \boldsymbol{a}_3^\top \mathbf{a}_2 = 0$ is the eigenvector with the third largest eigenvalue. As a result, $\boldsymbol{a}_3$ is the eigenvector with the third largest eigenvalue.