Week_1_mlt_questions

Jan 2024 Quiz 1

1. Suppose a dataset lies in $\mathbb{R}^4$ and undergoes Principal Component Analysis (PCA) after being centered. The resulting first and second principal components are given by:

$$\frac{1}{\sqrt{3}}\begin{bmatrix}1\\-1\\0\\1\end{bmatrix}, \quad \frac{1}{\sqrt{3}}\begin{bmatrix}1\\0\\1\\-1\end{bmatrix}$$

Which of the following could be the third principal component?

(a) $\frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\\0\\0\end{bmatrix}$

(b) $\frac{1}{\sqrt{3}}\begin{bmatrix}1\\1\\1\\0\end{bmatrix}$

(c) $\frac{1}{\sqrt{2}}\begin{bmatrix}-1\\-1\\0\\0\end{bmatrix}$

(d) $\frac{1}{\sqrt{3}}\begin{bmatrix}0\\1\\1\\1\end{bmatrix}$

[Ans: d]

2. **Comprehension:**
Standard PCA has been performed on a centered dataset in $\mathbb{R}^3$. The first two principal components are given below:

$$\mathbf{w}_1 = \frac{1}{\sqrt{3}}\begin{bmatrix}1\\1\\1\end{bmatrix}, \quad \mathbf{w}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\-1\\0\end{bmatrix}$$

Consider the data point in the dataset: $[2 \quad 1 \quad -1]^T$. (a, b) is the representation of this point in the coordinate system formed by the two principal components given above. The first and second coordinates correspond to PC-1 and PC-2 respectively. Based on the above data, answer the given subquestions.

i. What is the value of a? Enter your answer correctly to three decimal places.

[Ans: 1.10 to 1.20]

ii. What is the value of b? Enter your answer correctly to two decimal places.

[Ans: 0.70 to 0.72]

May 2024 Quiz 1

3. Let $\mathbf{w}_1$ and $\mathbf{w}_2$ be the top two principal components of the covariance matrix of a centered dataset. Let $\mathbf{x} \in \mathbb{R}^d$ be a data-point. Which of the following is the reconstruction error of the data-point after projecting it onto the top two principal components?

(a) $\|\mathbf{x} - (\mathbf{x}^T\mathbf{w}_1)\mathbf{w}_1 - (\mathbf{x}^T\mathbf{w}_2)\mathbf{w}_2\|^2$

(b) $\|\mathbf{x} - (\mathbf{x}^T\mathbf{w}_1)\mathbf{w}_1\|^2 + \|\mathbf{x} - (\mathbf{x}^T\mathbf{w}_2)\mathbf{w}_2\|^2$

(c) $(\mathbf{x}^T\mathbf{w}_1)^2 + (\mathbf{x}^T\mathbf{w}_2)^2$

(d) $\|\mathbf{x} - \mathbf{w}_1\|^2 + \|\mathbf{x} - \mathbf{w}_2\|^2$

[Ans: a]

4. Consider a covariance matrix C for a mean-centered dataset in $\mathbb{R}^3$. After performing standard PCA, the three principal components turn out to be $\begin{bmatrix}1\\0\\0\end{bmatrix}$, $\begin{bmatrix}0\\1\\0\end{bmatrix}$, $\begin{bmatrix}0\\0\\1\end{bmatrix}$. Which of the following statements are true? You can assume that C is not the zero matrix. (MSQ)

(a) C is a diagonal matrix.

(b) The diagonal entries of C are non-negative.

(c) The diagonal entries of C are strictly greater than zero.

(d) C has to be the identity matrix.

(e) C is a matrix of the form $k\mathbf{I}$, where $k > 0$ and $\mathbf{I}$ is the identity matrix.

[Ans: a, b]

5. **Comprehension:**
Let C be the covariance matrix of a mean-centered dataset:

$$C = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

Standard PCA is performed on this dataset. The first two PCs are given below:

$$\mathbf{w}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \mathbf{w}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

Based on the above data, answer the given subquestions.

i. Find the variance of the dataset along the first principal component.

[Ans: 3]

ii. Find the variance of the dataset along the x-axis.

[Ans: 2]

iii. Find the coordinates of the point $\begin{bmatrix} 2 \\ 2 \end{bmatrix}$ in the new coordinate system formed by the principal components.

(a) $\begin{bmatrix} 2\sqrt{2} \\ 0 \end{bmatrix}$

(b) $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$

(c) $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$

(d) $\begin{bmatrix} 2 \\ 0 \end{bmatrix}$

[Ans: a]

Sep 2024 Quiz 1

6. The eigenvalues of the covariance matrix of a centered dataset in $\mathbb{R}^5$ are $30, 10, 10, 0, 0$. Standard PCA is performed on this dataset. What is the variance captured by the top two principal components expressed as a percentage of total variance?

[Ans: 80]

7. Which of the following expressions is the reconstruction error for a dataset of n points, with respect to a line passing through the origin represented by the vector $\mathbf{w}$. Note that $\|\mathbf{w}\| = 1$. (MSQ)

(a) $\frac{1}{n} \sum_{i=1}^{n} \|\mathbf{x}_i - (\mathbf{x}_i^T \mathbf{w})\mathbf{w}\|^2$

(b) $\frac{1}{n} \sum_{i=1}^{n} [\mathbf{x}_i - (\mathbf{x}_i^T \mathbf{w})\mathbf{w}]^T [\mathbf{x}_i - (\mathbf{x}_i^T \mathbf{w})\mathbf{w}]$

(c) $\frac{1}{n} \sum_{i=1}^{n} [\mathbf{x}_i^T \mathbf{x}_i - (\mathbf{x}_i^T \mathbf{w})^2]$

(d) $-\frac{1}{n} \sum_{i=1}^{n} (\mathbf{x}_i^T \mathbf{w})^2$

[Ans: a, b]

8. **Comprehension:**

Given the vector $\mathbf{x} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$ and the line passing through the origin represented by the vector $\mathbf{w} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Answer the given subquestions.

i. Find the length of the projection of $\mathbf{x}$ onto the line defined by $\mathbf{w}$. Enter your answer correct to two decimal places.

[Ans: 4.8 to 5.1]

ii. Calculate the magnitude(norm) of reconstruction error after projecting $\mathbf{x}$ onto the line defined by $\mathbf{w}$. Enter your answer correct to two decimal places.

[Ans: 0.5 to 0.9]

Jan 2025 Quiz 1

9. Let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ be d-dimensional data points ($d > n$) and X be the matrix of shape $d \times n$ containing the data points. The k^{th} largest eigenvalue and corresponding unit eigenvector of $X^T X$ is λ and $\boldsymbol{\alpha}_k$, respectively. What will be the projection of $\mathbf{x}_i$ on the k^{th} principal component?

(a) $\mathbf{x}_i^T \boldsymbol{\alpha}_k$

(b) $\frac{\mathbf{x}_i^T \boldsymbol{\alpha}_k}{\sqrt{n\lambda}}$

(c) $\frac{\mathbf{x}_i^T X \boldsymbol{\alpha}_k}{\sqrt{\lambda}}$

(d) $\mathbf{x}_i^T X \boldsymbol{\alpha}_k$

[Ans: c]

10. Which of the following matrices are appropriate matrix $K = X^T X$ for some data matrix X? Choose the most appropriate answer. (MSQ)

(a) $K = \begin{bmatrix} 4 & 1 & 2 \\ 1 & 5 & 3 \\ 2 & 3 & 6 \end{bmatrix}$

(b) $K = \begin{bmatrix} -2 & -1 & 0 \\ -1 & -3 & -1 \\ 0 & -1 & -1 \end{bmatrix}$

(c) $K = \begin{bmatrix} 1 & -2 & 1 \\ 2 & 5 & -2 \\ 1 & -2 & 2 \end{bmatrix}$

(d) $K = \begin{bmatrix} 3 & 1 & 0 \\ 1 & 4 & 1 \\ 0 & 1 & 5 \end{bmatrix}$

[Ans: a, d]

11. Let X be a data matrix of shape (d, n) for a centered dataset. The first principal component of the dataset is given as

$$\left[\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right]^T$$

What will be the scalar proxy of the point $\begin{bmatrix}3\\4\end{bmatrix}$ onto the first principal component? Enter the answer correct to two decimal places.

[Ans: 4.92 to 4.98]

12. Consider a dataset that has 1000 samples, where each sample belongs to $\mathbb{R}^{20}$. PCA is run on this dataset and the top 5 principal components are retained, the rest being discarded. If it takes one unit of memory to store a real number, find the percentage decrease in storage space of the dataset by moving to its compressed representation. Enter your answer correct to two decimal places.

[Ans: 74 to 75]

13. **Comprehension:**
Consider the following dataset consisting of four points, all of which are collinear:

$$S = \left\{ \begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1\\2\end{bmatrix}, \begin{bmatrix}3\\6\end{bmatrix}, \begin{bmatrix}5\\10\end{bmatrix} \right\}$$

Based on the above data, answer the given subquestions.

i. Among the vectors given below, choose a representative that has unit length.

(a) $\begin{bmatrix}1\\3\\2\\3\end{bmatrix}$

(b) $\begin{bmatrix}\frac{1}{\sqrt{5}}\\\frac{2}{\sqrt{5}}\end{bmatrix}$

(c) $\begin{bmatrix}\frac{3}{\sqrt{5}}\\\frac{6}{\sqrt{5}}\end{bmatrix}$

(d) $\begin{bmatrix}\frac{3}{5}\\\frac{4}{5}\end{bmatrix}$

[Ans: b]

ii. If standard PCA is performed on this dataset, what is the variance along the first principal component? Enter the answer correct to two decimal places.

[Ans: 43 to 44]

May 2025 Quiz 1

14. Consider a dataset X of shape (d, n), with $d > n$. The k-th principal component of the dataset can be written as $\mathbf{w}_k = \sum_{i=1}^{n} \alpha_i \mathbf{x}_i$, where the vector $\mathbf{x}_i$ is the i-th data point. The k-th largest eigenvalue and the corresponding eigenvector of $X^T X$ are 10 and $\left[\frac{2}{\sqrt{54}}, \frac{3}{\sqrt{54}}, \frac{4}{\sqrt{54}}, \frac{6}{\sqrt{54}}\right]^T$, respectively. What will be the value of α_2? Enter the answer correct to two decimal places.

[Ans: 0.12 to 0.13]

15. Consider a dataset $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{50}\}$, where $\mathbf{x}_i \in \mathbb{R}^{1000}$. After performing PCA on the dataset, suppose that the residues become zero after the fifth round. Then, how many numbers do we need to store for this dataset?

[Ans: 5250]

16. **Comprehension:**
Consider the following data points:

$$\mathbf{x}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

Standard PCA is run on this dataset. Based on the above data, answer the given subquestions.

i. Which of the following represents the covariance matrix C?

(a) $\frac{1}{3}\begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix}$

(b) $\frac{1}{3}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$

(c) $\frac{1}{3}\begin{bmatrix} 0 & 0 & 0 \\ 0 & 5 & 4 \\ 0 & 4 & 5 \end{bmatrix}$

(d) $\frac{1}{3}\begin{bmatrix} 0 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix}$

[Ans: b]

ii. What is the variance of the dataset along the first principal component?

[Ans: 1]

iii. Find the residues after projecting the data points onto the first principal component. Note: Here $\mathbf{x}'_i$ represents the residues of the data point $\mathbf{x}_i$.

(a) $\mathbf{x}'_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \mathbf{x}'_2 = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}, \quad \mathbf{x}'_3 = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$

(b) $\mathbf{x}'_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \mathbf{x}'_2 = \begin{bmatrix} 0.5 \\ -0.5 \end{bmatrix}, \quad \mathbf{x}'_3 = \begin{bmatrix} -0.5 \\ 0.5 \end{bmatrix}$

(c) $\mathbf{x}'_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \mathbf{x}'_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad \mathbf{x}'_3 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$

(d) $\mathbf{x}'_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \mathbf{x}'_2 = \begin{bmatrix} 0 \\ -1 \end{bmatrix}, \quad \mathbf{x}'_3 = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$

[Ans: b]

2024 Apr 28 ET

17. Consider a dataset X in $\mathbb{R}^3$. The dataset X consists of 4 samples with 3 features each. The covariance matrix C of this dataset has three non-zero eigenvalues which follow the given linear equations:

$$2\lambda_1 + 3\lambda_2 - \lambda_3 = 5$$
$$\lambda_1 - 2\lambda_2 + 4\lambda_3 = 8$$
$$3\lambda_1 + \lambda_2 - 2\lambda_3 = 3$$

Determine the variance of the given dataset.

[Ans: 5]

2024 Sep 01 ET

18. The eigenvalues of the covariance matrix of a centered dataset in $\mathbb{R}^5$ are $15, 5, 5, 0, 0$. Standard PCA is performed on this dataset. What is the variance captured by the top two principal components expressed as a percentage of total variance?

(a) 80%

(b) 60%

(c) 20%

(d) 15%

[Ans: a]

2024 Dec 22 ET

19. Consider the following dataset of 4 points in $\mathbb{R}^5$:

$$\{-2\mathbf{u}, -\mathbf{u}, \mathbf{u}, 2\mathbf{u}\}$$

where $\mathbf{u} = [0.5 \quad 0 \quad 0.5 \quad -0.5 \quad -0.5]^T$. What is the variance along the first principal component after performing standard PCA on this dataset? Enter the answer correct to one decimal place.

[Ans: 2.5]

20. Consider a dataset of 200 points where the i-th data point is given by:

$$\mathbf{x}_i = a_i \begin{bmatrix} 3 \\ 1 \\ 0 \\ 2 \end{bmatrix} + b_i \begin{bmatrix} 1 \\ 0 \\ 3 \\ 0 \end{bmatrix},$$

where a_i and b_i are real numbers such that

$$\sum_{i=1}^{200} a_i = \sum_{i=1}^{200} b_i = 0.$$

Standard PCA is performed on this dataset. If the top two principal components are retained and used to reconstruct the dataset, what is the reconstruction error?

[Ans: 0]

2025 Aug 31 ET

21. Consider a dataset $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4\}$, where each $\mathbf{x}_i \in \mathbb{R}^3$. The covariance matrix C has three non-zero eigenvalues $\lambda_1 > \lambda_2 > \lambda_3$ satisfying:

$$2\lambda_1 - \lambda_2 + \lambda_3 = 8$$
$$\lambda_1 + 2\lambda_2 - \lambda_3 = 5$$

The difference between the largest and smallest eigenvalue is 3. Determine the total variance of the dataset.

[Ans: 6]

2025 Dec 18 ET

22. A centered dataset in $\mathbb{R}^4$ is projected onto its first two principal components. The first and the second principal components are respectively:

$$v_1 = \frac{1}{2}\begin{bmatrix}1\\1\\1\\1\end{bmatrix}, \quad v_2 = \frac{1}{2}\begin{bmatrix}1\\-1\\1\\-1\end{bmatrix}$$

Then, which of the following vectors could be a valid third principal component? (MSQ)

(a) $\frac{1}{2}\begin{bmatrix}1\\1\\-1\\-1\end{bmatrix}$

(b) $\frac{1}{2}\begin{bmatrix}1\\-1\\-1\\1\end{bmatrix}$

(c) $\frac{1}{4}\begin{bmatrix}2\\2\\2\\2\end{bmatrix}$

(d) $\begin{bmatrix}0\\0\\0\\0\end{bmatrix}$

[Ans: a, b]

2026 May 06 ET

23. Does PCA reduce dimensionality if all 5 principal components of a dataset in $\mathbb{R}^5$ are retained?

(a) Yes, because PCA eliminates redundant information regardless of the number of components retained.

(b) No, because retaining all 5 principal components preserves the full 5-dimensional representation (performing only a rotation/change of basis).

(c) Yes, provided the covariance matrix is diagonal.

(d) It depends on whether the dataset is mean-centered.

[Ans: b]

2026 Sep 13 ET

24. For the covariance matrix

$$\Sigma = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix},$$

compute the variance captured along the first principal component.

(a) 2

(b) 3

(c) 4

(d) 6

[Ans: c]

Oct 2022 Quiz 1

25. Let X be a data matrix of shape (d, n) for a centered dataset. The first principal component of the dataset is $\left[\frac{\sqrt{5}}{2}, \frac{2}{3}\right]^T$. What will the scalar proxy of the point $[2, 3]^T$ be on the first principal component?

(a) $\frac{2\sqrt{5}}{3} + 6$

(b) $\frac{2\sqrt{5}+2}{3}$

(c) $\frac{2\sqrt{5}}{9} + 2$

(d) $\frac{2\sqrt{5}}{3} + 2$

[Ans: d]

26. Consider a centered dataset of 100 points in $\mathbb{R}^{10}$. Standard PCA is performed on this dataset. Consider the following ratio:

$$\theta = \frac{\text{variance along the first principal component}}{\text{sum of variances along all ten principal components}}.$$

What is the minimum value of θ?

[Ans: 0.1]

27. Consider a centered dataset of 1000 points in $\mathbb{R}^{10}$. Standard PCA is performed on this dataset. What is the maximum number of principal components that can be retained so that the compression ratio does not fall below 1.64? The compression ratio is defined as:

$$\frac{\text{size of original dataset}}{\text{size of reconstructed dataset}}.$$

Assume that each real number occupies one unit of storage space. Note that we are interested in the reconstructions and not just in the scalar projections. So, the reconstructed dataset will also be in $\mathbb{R}^{10}$.

[Ans: 6]

28. **Comprehension:**
Consider a dataset that has 4 points in $\mathbb{R}^2$ that lie on a line passing through the origin:

$$D = \left\{ \begin{bmatrix} -4 \\ -2 \end{bmatrix}, \begin{bmatrix} -2 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 4 \\ 2 \end{bmatrix} \right\}.$$

Based on the above data, answer the given subquestions.

i. Is this dataset centered?

(a) Yes

(b) No

[Ans: a]

ii. Choose all representatives $\mathbf{w}$ for this dataset such that $\|\mathbf{w}\| = 1$. (MSQ)

(a) $\frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ 1 \end{bmatrix}$

(b) $\frac{1}{\sqrt{5}} \begin{bmatrix} -2 \\ -1 \end{bmatrix}$

(c) $\begin{bmatrix} -2 \\ -1 \end{bmatrix}$

(d) $\begin{bmatrix} 4 \\ 2 \end{bmatrix}$

[Ans: a, b]

iii. If standard PCA is performed on this dataset, what is the variance along the first principal component?

[Ans: 12.5]

iv. If standard PCA is performed on this dataset, what is the variance along the second principal component?

[Ans: 0]

Jul 2023 Quiz 1

29. Select all true statements regarding standard PCA applied on a centered dataset. (MSQ)

(a) The variance of the dataset along the first principal component is maximum.

(b) The variance of the dataset along the first principal component is minimum.

(c) The first principal component is the unit norm eigenvector corresponding to the largest eigenvalue of the covariance matrix.

(d) The first principal component is the unit norm eigenvector corresponding to the smallest eigenvalue of the covariance matrix.

[Ans: a, c]

30. Consider a dataset of 2000 points all of which lie in $\mathbb{R}^{30}$. If all these points lie in a four-dimensional low subspace then after applying the PCA algorithm which of the following statements is/are true? (MSQ)

(a) The covariance matrix corresponding to this dataset has 30 non-zero eigenvalues.

(b) The covariance matrix corresponding to this dataset has only four non-zero eigenvalues.

(c) The residues will become zero after four rounds.

(d) By utilizing four principal component vectors and their corresponding coefficients, we have the ability to reconstruct the complete dataset using representatives.

[Ans: b, c, d]

31. **Comprehension:**
Standard PCA has been performed on a centered dataset in $\mathbb{R}^3$. The first two principal components are given below:

$$\mathbf{w}_1 = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \quad \mathbf{w}_2 = \frac{1}{\sqrt{6}} \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}.$$

Consider the following data point in the dataset: $[1 \quad 2 \quad 1]^T$. (a, b) is the representation of this point in the coordinate system formed by the two principal components given above. The first and second coordinates correspond to PC-1 and PC-2 respectively. Based on the above data, answer the given subquestions.

i. What is the value of a? Enter your answer correct to three decimal places.

[Ans: 1.342]

ii. What is the value of b? Enter your answer correct to two decimal places.

[Ans: 0.41]

Oct 2023 Quiz 1

32. Which of the following statements are true regarding applying standard PCA on a centered dataset? (MSQ)

(a) The variance of the dataset along the first principal component is maximum.

(b) Eigenvectors are always orthogonal to each other in PCA.

(c) The first principal component is the eigenvector corresponding to the largest eigenvalue of the covariance matrix.

(d) The first principal component is the eigenvector corresponding to the smallest eigenvalue of the covariance matrix.

[Ans: a, b, c]

33. Consider a dataset that has 4 points in $\mathbb{R}^2$ that lie on a line passing through the origin:

$$D = \left\{ \begin{bmatrix} -10 \\ -30 \end{bmatrix}, \begin{bmatrix} -5 \\ -15 \end{bmatrix}, \begin{bmatrix} 5 \\ 15 \end{bmatrix}, \begin{bmatrix} 10 \\ 30 \end{bmatrix} \right\}.$$

If normal PCA is performed on this dataset, what is the variance along the first principal component?

[Ans: 625]

34. Consider a dataset of 50 points all of which lie in $\mathbb{R}^5$. The eigenvalues of the covariance matrix are given below:

$$1.7, \quad 1.4, \quad 0.25, \quad 0.20, \quad 0.005.$$

If we run the PCA algorithm on this dataset, and retain the top-k principal components, what is a good choice of k? (Hint: A good choice means it captures at least 95% of variance.)

[Ans: 4]

Oct 2025 Quiz 1

35. Consider a dataset consisting of 5000 points in $\mathbb{R}^{50}$. If all these points lie in a 6-dimensional linear subspace, then which of the following will be true after applying PCA? (MSQ)

(a) There will be 50 non-zero eigenvalues of the covariance matrix.

(b) There will be 6 non-zero eigenvalues of the covariance matrix.

(c) Using the first 5 principal components, the dataset can be reconstructed with zero residue.

(d) The variance captured by the top 6 principal components equals the total variance of the dataset.

[Ans: b, d]

36. Consider a mean-centered dataset consisting of 150 patients from the health-care domain with 6 features. The dataset is represented as a 150×6 matrix X. PCA is performed on this dataset, and it is observed that the reconstruction error (residue) vanishes completely after k components. A domain expert provides the following insights about the columns of X, denoted by $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_6$:

(a) The set $\{\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3\}$ is linearly independent.

(b) The following relationships hold:

$$\mathbf{c}_4 = 2\mathbf{c}_1 - \mathbf{c}_2, \quad \mathbf{c}_5 = \mathbf{c}_2 + \mathbf{c}_3, \quad \mathbf{c}_6 = 4\mathbf{c}_1 - 3\mathbf{c}_3.$$

What is the value of k?

[Ans: 3]

37. Consider a dataset $X \in \mathbb{R}^{d \times n}$ with sample size $n = 100$ (i.e., X has d features and 100 examples), where $d \gg 100$. Suppose the first five eigenvalues of the matrix $X^T X$ are:

$$10, \quad 5, \quad 3, \quad 2, \quad 1,$$

and all remaining eigenvalues are zero. What is the proportion of variance explained along the first principal component? Enter your answer correct to two decimal places.

[Ans: 0.47]

38. **Comprehension:**
Consider the following 2D dataset shown in Figure 1. PCA is to be applied on this dataset. Based on the above data, answer the given subquestions.

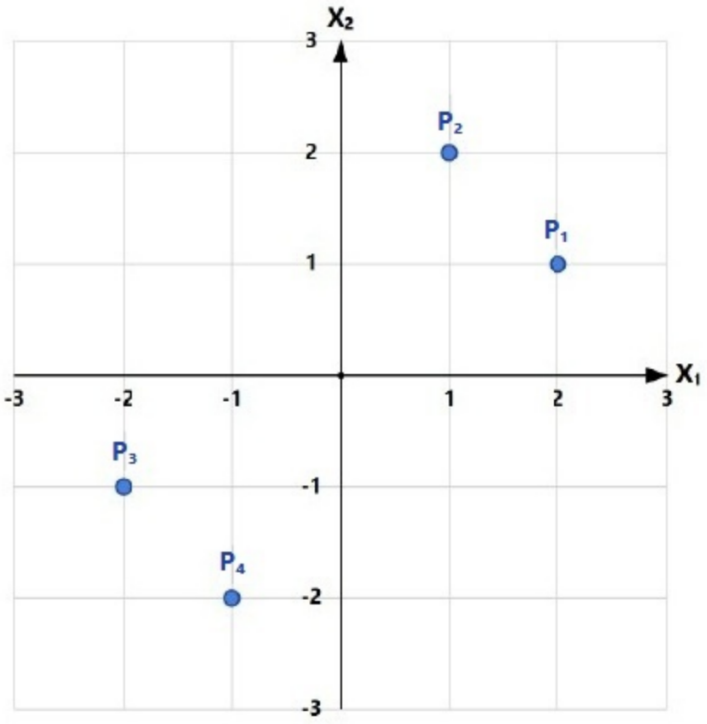


Figure 1: Dataset 1

i. What is the covariance matrix of the dataset?

(a) $\begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix}$

(b) $\begin{bmatrix} 2.5 & 2 \\ 2 & 2.5 \end{bmatrix}$

(c) $\begin{bmatrix} 3.5 & 2.5 \\ 2.5 & 3.5 \end{bmatrix}$

(d) $\begin{bmatrix} 4 & 3 \\ 3 & 4 \end{bmatrix}$

[Ans: b]

ii. If standard PCA is performed on this dataset, what is the variance along the first principal component? Enter the answer correct to one decimal place.

[Ans: 4.5]

iii. If standard PCA is performed on this dataset, what is the variance along the second principal component? Enter the answer correct to one decimal place.

[Ans: 0.5]

Mar 2026 Quiz 1

39. Consider a centered dataset D in $\mathbb{R}^4$ such that every data point (x_1, x_2, x_3, x_4) in D satisfies the equation

$$2x_1 + x_2 - x_3 + 2x_4 = 0.$$

(a) $\frac{1}{\sqrt{10}}(2, 1, -1, 2)$

(b) $\frac{1}{\sqrt{10}}(-2, -1, 1, -2)$

(c) $\frac{1}{\sqrt{6}}(1, 0, 2, 1)$

(d) $\frac{1}{3}(2, 0, 1, 2)$

[Ans: a, b]

Source note: This is an MSQ, but the selection criterion is missing in the supplied reference.

40. **Comprehension:**
Consider a cube of edge-length 2 centered at the origin with its corners represented as

$$(\pm 1, \pm 1, \pm 1).$$

These eight points make up a dataset D. Based on the above data, answer the given subquestions.

i. Find the covariance matrix C of the dataset. Enter the sum of all the entries of C as the answer.

[Ans: 3]

ii. Let $\mathbf{v} \in \mathbb{R}^3$ be some unit vector. Find the variance of the dataset along the direction $\mathbf{v}$.

[Ans: 1]

41. **Comprehension:**
Standard PCA is performed on a centered dataset in $\mathbb{R}^3$. The first three principal components are:

$$\mathbf{w}_1 = \frac{1}{\sqrt{2}}(1, 1, 0), \quad \mathbf{w}_2 = \frac{1}{\sqrt{2}}(-1, 1, 0), \quad \mathbf{w}_3 = (0, 0, 1).$$

The first two principal components are retained and the third is discarded. The coordinate representation of the data point $(1, 2, 3)$ after projection onto the first two principal components is (a, b). Based on the above data, answer the given subquestions.

i. Find a. Enter your answer correct to three decimal places.

[Ans: 2.121]

ii. Find b. Enter your answer correct to three decimal places.

[Ans: 0.707]

Jul 2026 Quiz 1

42. Consider a mean-centered dataset D consisting of 5000 points in $\mathbb{R}^{10}$ with features $x_1, x_2, \ldots, x_{10}$. Every data point satisfies $x_1 = x_2 = x_3$, $x_9 = 0$, $x_{10} = 0$. The independent base feature x_1 has variance 100. The remaining features x_4, x_5, x_6, x_7, x_8 are mutually uncorrelated, uncorrelated with x_1, and each has variance 10. Standard PCA is performed on this dataset. Let

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{10}$$

denote the eigenvalues of the covariance matrix in descending order. Which of the following statements is strictly true?

(a) Exactly 3 eigenvalues are zero: $\lambda_8 = \lambda_9 = \lambda_{10} = 0$, and $\text{PC}_1 = (1, 0, 0, 0, 0, 0, 0, 0, 0, 0)$.

(b) Exactly 4 eigenvalues are zero: $\lambda_7 = \lambda_8 = \lambda_9 = \lambda_{10} = 0$, and $\text{PC}_1 = \frac{1}{\sqrt{3}}(1, 1, 1, 0, 0, 0, 0, 0, 0, 0)$.

(c) Exactly 4 eigenvalues are zero: $\lambda_7 = \lambda_8 = \lambda_9 = \lambda_{10} = 0$, and $\text{PC}_1 = (1, 0, 0, 0, 0, 0, 0, 0, 0, 0)$.

(d) Exactly 2 eigenvalues are zero: $\lambda_9 = \lambda_{10} = 0$, and PC_1 lies somewhere in the subspace spanned by x_1, x_2, x_3.

[Ans: b]

43. Let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n \in \mathbb{R}^d$. Define

$$\mathbf{v}^* = \arg\min_{\mathbf{v} \in \mathbb{R}^d} \sum_{i=1}^{n} \|3\mathbf{x}_i - 2\mathbf{v}\|^2.$$

Which of the following is the correct expression for $\mathbf{v}^*$?

(a) $\mathbf{v}^* = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i$

(b) $\mathbf{v}^* = \frac{3}{2} \cdot \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i$

(c) $\mathbf{v}^* = \frac{3}{n} \sum_{i=1}^{n} \mathbf{x}_i$

(d) $\mathbf{v}^* = \frac{2}{3} \cdot \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i$

[Ans: b]

44. Let $D = \{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}\}$ be a dataset of n observations where each $\mathbf{x}^{(i)} \in \mathbb{R}^{100}$. It is given that

$$\sum_{i=1}^{n} \mathbf{x}^{(i)} = 0.$$

The covariance matrix computed from D has eigenvalues $\lambda_i = 100\,2^{-i}$, $1 \leq i \leq 100$. Let $\mathbf{u} \in \mathbb{R}^{100}$ be the direction of maximum variance with $\mathbf{u}^T\mathbf{u} = 1$. Find the value of

$$\frac{1}{n}\sum_{i=1}^{n}(\mathbf{u}^T\mathbf{x}^{(i)})^2.$$

[Ans: 100]

45. If the function is called as `compute_metric(45, 75)`, what value will the function return?

[Ans: 15]

Source note: The function definition is missing in the supplied reference.

Dec 2022 End Term

46. **Comprehension:**
Upon performing standard PCA on a centered dataset in $\mathbb{R}^3$, we get the principal components to be:

$$\mathbf{w}_1 = [1 \quad 0 \quad 0]^T, \quad \mathbf{w}_2 = [0 \quad 1 \quad 0]^T, \quad \mathbf{w}_3 = [0 \quad 0 \quad 1]^T.$$

C is the covariance matrix of the centered dataset. The off-diagonal entries are hidden from your view:

$$C = \begin{bmatrix} 15 & a & b \\ a & 9 & c \\ b & c & 3 \end{bmatrix}.$$

$[x_1 \quad x_2 \quad x_3]^T$ denotes a data point. Here, x_1, x_2, x_3 are the three features. Note: The word standard indicates that no kernel has been used. Based on the above data, answer the given subquestions.

i. What is the variance along the first principal component?

[Ans: 15]

ii. The off-diagonal elements of C correspond to the covariance between a pair of features (x_i, x_j) with $i \neq j$. Which of the following statements about the features of this dataset is true?

   (a) Each pair of features has a strong positive correlation. For instance, if x_1 increases, then x_2 also increases.

   (b) Each pair of features has a strong negative correlation. For instance, if x_1 increases, then x_2 decreases.

   (c) Each pair of features is uncorrelated. For instance, if x_1 increases, we can say nothing about the trend of x_2.

   (d) Unless we know the exact values of the off-diagonal elements a, b, c, we cannot comment about the correlation among features.

[Ans: c]

Apr 2023 End Term

47. Standard PCA was applied on a centered dataset in $\mathbb{R}^6$. All the eigenvalues of the covariance matrix of this centered dataset are given below in decreasing order with $\lambda > 1$:

$$\lambda^5, \quad \lambda^4, \quad \lambda^3, \quad \lambda^2, \quad \lambda, \quad 1.$$

If the sum of the variances along the first principal components is exactly 88.88% of the total variance, what is the value of λ? Enter the nearest integer as your answer. Note that the total variance is the sum of the variances along each principal component. Hint: $\lambda^2 + \lambda + 1$.

[Ans: 2]

Source note: The number of retained principal components is not specified in the supplied reference.

Sep 2023 End Term

48. When PCA is applied to a certain mean-centered dataset in $\mathbb{R}^5$, the eigenvalues of the covariance matrix are $10, 15, 9, 0, 0$ and the corresponding eigenvectors are

$$\left[\sqrt{\frac{1}{2}}, \sqrt{\frac{1}{2}}, 0, 0, 0\right]^T, \quad \left[-\sqrt{\frac{1}{2}}, \sqrt{\frac{1}{2}}, 0, 0, 0\right]^T, \quad [0, 0, 1, 0, 0]^T,$$
$$\left[0, 0, 0, \sqrt{\frac{1}{2}}, -\sqrt{\frac{1}{2}}\right]^T, \quad \left[0, 0, 0, -\sqrt{\frac{1}{2}}, \sqrt{\frac{1}{2}}\right]^T.$$

What is the variance of the dataset?

[Ans: 34]

Dec 2023 End Term

49. Consider the dataset $D = \{(-1,-1),(0,1),(1,1)\}$. What is the first principal component (i.e., the direction corresponding to the largest eigenvalue of the covariance matrix) for the above dataset? (MSQ)

(a) $\begin{bmatrix}1\\0\end{bmatrix}$

(b) $\begin{bmatrix}0\\1\end{bmatrix}$

(c) $\begin{bmatrix}1\\1\end{bmatrix}$

(d) $\begin{bmatrix}-1\\0\end{bmatrix}$

[Ans: a, d]

50. A team was given a dataset $X \in \mathbb{R}^{d\times n}$ where d denotes the number of features and n denotes the number of samples. They found that there are 10 samples in the dataset and each sample contains 100 features. Assume that the data points $\mathbf{x}_4$ to $\mathbf{x}_{10}$ are all linear combinations of the linearly independent data points $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$.

Suppose the team applies linear PCA on the dataset and reconstructs the data points with zero error using k principal components (directions). For which value of k would the reconstruction error become zero?

(a) 1

(b) 2

(c) 3

(d) 5

(e) 10

(f) 100

[Ans: c]