DAI-101 Tutorial – 1 (5 February)

Instructions:

- Total Time: 25 minutes
- Total Marks: 20 marks
- Attempt all questions
- No negative marking

[1 $\times$ 10 = 10 marks]

1. What is the correct interpretation of the standard deviation of a dataset?

(A) It is the square of the mean
(B) It measures the asymmetry of the distribution
(C) It represents the spread of data around the mean
(D) It is always equal to the variance

2. According to Chebyshev's inequality, for any dataset (not necessarily normal), at least what percentage of values lie within $k = 2$ standard deviations from the mean?

(A) 50%
(B) 75%
(C) 88.89%
(D) 95%

3. Which of the following matrices is always symmetric?

(A) Covariance matrix
(B) Confusion matrix
(C) Min-max scaled matrix
(D) One-hot encoded matrix

4. Which of the following measures is most affected by extreme values (outliers)?

(A) Mean
(B) Median
(C) Mode
(D) Interquartile range

5. The eigenvectors of a real symmetric matrix are:

(A) Always real and orthogonal
(B) Always complex

(C) Always linearly dependent

(D) Not guaranteed to exist

6. Which of the following properties is true for a matrix to be positive semi-definite?

(A) All eigenvalues are negative

(B) All eigenvalues are non-negative

(C) It must be a diagonal matrix

(D) It must be orthogonal

7. The trace of a matrix is equal to:

(A) The sum of all elements in the matrix

(B) The sum of all eigenvalues

(C) The product of the diagonal elements

(D) The determinant of the matrix

8. A matrix is orthogonal if:

(A) Its determinant is 0

(B) Its inverse is equal to its transpose

(C) All rows are equal

(D) It has complex eigenvalues

9. In the context of PCA, the eigenvector corresponding to the largest eigenvalue represents:

(A) The direction of minimum variance

(B) A redundant feature

(C) The principal component with the most variance

(D) The mean of the dataset

10. Which of the following is not a valid reason for using PCA?

(A) Dimensionality reduction

(B) Noise reduction

(C) Improving classification accuracy directly

(D) Visualizing high-dimensional data

[2 marks]
Q11. What is the geometric intuition behind Principal Component Analysis (PCA)? Explain why the first principal component is important.

[4 marks]
Q12. Perform the spectral decomposition of the following symmetric matrix:

$$A = \begin{pmatrix} 4 & 1 \\ 1 & 3 \end{pmatrix}$$

[2 marks]
Q13. Given a NumPy array X of shape (n_samples, n_features), center the data and compute the covariance matrix manually (do not use `np.cov`).

```
import numpy as np

X = np.array([[1, 2], [3, 4], [5, 6]])

# TODO: Your code after this line
```

[2 marks]
Q14. Using pandas, load the following DataFrame and perform: Min-max scaling on "age", Z-score standardization on "income", and one-hot encoding on "city".

```
import pandas as pd

data = pd.DataFrame({
    'age': [22, 25, 47, 52],
    'income': [20000, 25000, 47000, 52000],
    'city': ['Delhi', 'Mumbai', 'Delhi', 'Chennai']
})

# TODO: Your code after this line
```

DAI-101 Autumn 2026
Tutorial-02

05 February 2026

Q1. MCQ: In the SVD factorization $A = U\Sigma V^T$, which of the following is true?

- (A) U and V are diagonal
- (B) The columns of U are orthonormal
- (C) Σ must be square
- (D) V is upper triangular

Q2. MCQ: What are the singular values of the diagonal matrix

$$D = \begin{pmatrix} 3 & 0 \\ 0 & 5 \end{pmatrix}?$$

- (A) $3, 5$
- (B) $5, 3$
- (C) $\sqrt{3}, \sqrt{5}$
- (D) $3, 0, 0, 5$

Q3. Numerical: Suppose the eigenvalues of a covariance matrix (sorted descending) are:

$$\lambda_1 = 4.5,\ \lambda_2 = 2.0,\ \lambda_3 = 1.0,\ \lambda_4 = 0.5.$$

What is the minimum number of principal components needed to explain at least 90% of the variance?

Q4. Numerical: A centred data matrix $X \in \mathbb{R}^{100\times 50}$ has rank 40. Keeping the first $k = 20$ singular values yields

$$\sum_{i=1}^{20} \sigma_i^2 = 8000, \qquad \sum_{i=1}^{40} \sigma_i^2 = 9600.$$

Compute the fraction of total variance retained.

Q5. True/False: If you keep all non-zero singular values of X in its SVD, the reconstruction error (in Frobenius norm) is zero.

Q6. MCQ: In PCA, we compute the covariance matrix C. Which of the following is not a property of C?

- (A) C is symmetric
- (B) All eigenvalues of C are real
- (C) C is orthogonal
- (D) C is positive semi-definite

Q7. MCQ: Given an $m \times n$ matrix A with SVD $A = U\Sigma V^T$, which holds?

(A) Columns of U are eigenvectors of A

(B) Columns of V are eigenvectors of $A^T A$

(C) Σ is an orthogonal matrix

(D) U and V are diagonal matrices

Q8. Numerical: For

$$X = \begin{pmatrix} 2 & 0 \\ 0 & 2 \\ 3 & 1 \end{pmatrix},$$

(a) Compute the mean.

Q9. MCQ: If eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$, then variance retained by first k PCs is

(A) $\frac{\sum_{i=1}^k \lambda_i}{n}$

(B) $\frac{\sum_{i=1}^k \lambda_i}{\sum_{j=1}^n \lambda_j}$

(C) $\sum_{i=1}^k \lambda_i$

(D) λ_k

Q10. Numerical: Given $\sigma_1 = 5,\ \sigma_2 = 3,\ \sigma_3 = 1$, what fraction of total energy is in first two?

Q11. MCQ: Which of the following is **NOT** an application of Singular Value Decomposition (SVD)?

(A) Image compression via low-rank approximation

(B) Collaborative filtering in recommender systems

(C) Topic modeling in text analysis (Latent Semantic Analysis)

(D) Efficient sorting of numerical arrays

Q12. Coding: Compute the SVD of a simple 2×2 matrix and reconstruct it.

(1) Let

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}.$$

(2) Compute its SVD: $A = U\Sigma V^T$.

(3) Reconstruct $\hat{A} = U\Sigma V^T$ and print the Frobenius-norm error $\|A - \hat{A}\|_F$.

Q13. Coding: Use scikit-learn's PCA to reduce a small random dataset and report explained variance.

(1) Generate a random array X of shape $(10, 3)$.

(2) Fit `PCA(n_components=2)` on X and print explained variance ratio.

Starter Code (Optional)
Q12 (NumPy SVD):

```python
import numpy as np

A = np.array([[1, 2],
              [3, 4]], dtype=float)

U, s, VT = np.linalg.svd(A, full_matrices=False)
Sigma = np.diag(s)

A_hat = U @ Sigma @ VT
err = np.linalg.norm(A - A_hat, ord='fro')

print("U=\n", U)
print("Sigma=\n", Sigma)
print("VT=\n", VT)
print("Frobenius error =", err)
```

Q13 (scikit-learn PCA):

```python
import numpy as np
from sklearn.decomposition import PCA

np.random.seed(0)
X = np.random.randn(10, 3)

pca = PCA(n_components=2)
Z = pca.fit_transform(X)

print("Explained variance ratio:", pca.explained_variance_ratio_)
```