# Mathematical Foundation of Machine Learning Assignment III

Shedrack Eze
Andrew ID: seze

## Question 1: Matrix Properties and Invertibility

We are given:

$$A = \begin{pmatrix} 1 & 2 & 0 \\ 2 & 4 & 1 \\ 2 & 1 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix}, \quad C = \begin{pmatrix} 4 & 2 & 2 \\ 2 & 2 & 0 \\ 2 & 0 & 2 \end{pmatrix}.$$

### (a) Compute $\det(A)$ and $\mathbf{tr}(A)$. Is $A$ invertible?

**Step 1: Compute trace**

$$\mathrm{tr}(A) = a_{11} + a_{22} + a_{33} = 1 + 4 + 1 = 6.$$

**Step 2: Next we calculate determinant using cofactor expansion along the first row**

$$\det(A) = (+)a_{11} \begin{bmatrix} 4 & 1 \\ 1 & 1 \end{bmatrix} - a_{12} \begin{bmatrix} 2 & 1 \\ 2 & 1 \end{bmatrix} + a_{13} \begin{bmatrix} 2 & 4 \\ 2 & 1 \end{bmatrix}.$$

Substitute $a_{11} = 1$, $a_{12} = 2$, $a_{13} = 0$:

$$\det(A) = (1) \begin{bmatrix} 4 & 1 \\ 1 & 1 \end{bmatrix} - (2) \begin{bmatrix} 2 & 1 \\ 2 & 1 \end{bmatrix} + (0) \begin{bmatrix} 2 & 4 \\ 2 & 1 \end{bmatrix}.$$

Compute first minor:

$$\begin{bmatrix} 4 & 1 \\ 1 & 1 \end{bmatrix} = (4)(1) - (1)(1) = 4 - 1 = 3.$$

Compute second minor:

$$\begin{bmatrix} 2 & 1 \\ 2 & 1 \end{bmatrix} = (2)(1) - (1)(2) = 2 - 2 = 0.$$

Compute third minor:

$$\begin{bmatrix} 2 & 4 \\ 2 & 1 \end{bmatrix} = (2)(1) - (4)(2) = 2 - 8 = -6.$$

Now substitute:

$$\det(A) = (1)(3) - (2)(0) + (0)(-6) = 3 - 0 + 0 = 3.$$

**Step 3: Invertibility** Since $\det(A) = 3 \neq 0$, $A$ is invertible.

## (b) Find all eigenvalues of $B$ and verify trace and determinant relationships

Matrix:

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

**Step 1: Compute characteristic polynomial**

$$\det(B - \lambda I) = \begin{bmatrix} 2-\lambda & 1 \\ 1 & 3-\lambda \end{bmatrix}.$$

Compute determinant:

$$(2-\lambda)(3-\lambda) - (1)(1).$$

Expand $(2-\lambda)(3-\lambda)$:

$$(2)(3) + (2)(-\lambda) + (-\lambda)(3) + (-\lambda)(-\lambda) = 6 - 2\lambda - 3\lambda + \lambda^2 = \lambda^2 - 5\lambda + 6.$$

Subtract 1:

$$\lambda^2 - 5\lambda + 6 - 1 = \lambda^2 - 5\lambda + 5.$$

So characteristic equation:

$$\lambda^2 - 5\lambda + 5 = 0.$$

**Step 2: Solve quadratic using formula**

$$\lambda = \frac{-(-5) \pm \sqrt{(-5)^2 - 4(1)(5)}}{2(1)} = \frac{5 \pm \sqrt{25-20}}{2} = \frac{5 \pm \sqrt{5}}{2}.$$

So:

$$\lambda_1 = \frac{5-\sqrt{5}}{2} \approx 1.382, \quad \lambda_2 = \frac{5+\sqrt{5}}{2} \approx 3.618.$$

**Step 3: Verify trace**

$$\lambda_1 + \lambda_2 = \frac{5-\sqrt{5}}{2} + \frac{5+\sqrt{5}}{2} = \frac{10}{2} = 5 = \mathrm{tr}(B).$$

**Step 4: verify determinant**

$$\lambda_1 \cdot \lambda_2 = \left(\frac{5-\sqrt{5}}{2}\right)\left(\frac{5+\sqrt{5}}{2}\right) = \frac{(5)^2 - (\sqrt{5})^2}{4} = \frac{25-5}{4} = \frac{20}{4} = 5 = \det(B).$$

## (c) Compute Cholesky decomposition of $C$

Matrix:

$$C = \begin{pmatrix} 4 & 2 & 2 \\ 2 & 2 & 0 \\ 2 & 0 & 2 \end{pmatrix}.$$

Assume $C = LL^T$ where:

$$L = \begin{pmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{pmatrix}.$$

**Step 1: Compute $L^T$**

First compute the transpose of $L$:

$$L^T = \begin{pmatrix} l_{11} & l_{21} & l_{31} \\ 0 & l_{22} & l_{32} \\ 0 & 0 & l_{33} \end{pmatrix} = \begin{pmatrix} l_{11} & l_{21} & l_{31} \\ 0 & l_{22} & l_{32} \\ 0 & 0 & l_{33} \end{pmatrix}$$

**Step 2: Compute $LL^T$**

Multiply $L$ and $L^T$:

$$LL^T = \begin{pmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{pmatrix} \begin{pmatrix} l_{11} & l_{21} & l_{31} \\ 0 & l_{22} & l_{32} \\ 0 & 0 & l_{33} \end{pmatrix}$$

Compute each element:

Element $R_{11}$:

$$= (l_{11} \times l_{11}) + (0 \times 0) + (0 \times 0) = l_{11}^2 + 0 + 0 = l_{11}^2$$

Element $R_{12}$:

$$= (l_{11} \times l_{21}) + (0 \times l_{22}) + (0 \times 0) = l_{11}l_{21} + 0 + 0 = l_{11}l_{21}$$

Element $R_{13}$:

$$= (l_{11} \times l_{31}) + (0 \times l_{32}) + (0 \times l_{33}) = l_{11}l_{31} + 0 + 0 = l_{11}l_{31}$$

Element $R_{21}$:

$$= (l_{21} \times l_{11}) + (l_{22} \times 0) + (0 \times 0) = l_{21}l_{11} + 0 + 0 = l_{21}l_{11}$$

Element $R_{22}$:

$$= (l_{21} \times l_{21}) + (l_{22} \times l_{22}) + (0 \times 0) = l_{21}^2 + l_{22}^2 + 0 = l_{21}^2 + l_{22}^2$$

Element $R_{23}$:

$$= (l_{21} \times l_{31}) + (l_{22} \times l_{32}) + (0 \times l_{33}) = l_{21}l_{31} + l_{22}l_{32} + 0 = l_{21}l_{31} + l_{22}l_{32}$$

Element $R_{31}$:

$$= (l_{31} \times l_{11}) + (l_{32} \times 0) + (l_{33} \times 0) = l_{31}l_{11} + 0 + 0 = l_{31}l_{11}$$

Element $R_{32}$:

$$= (l_{31} \times l_{21}) + (l_{32} \times l_{22}) + (l_{33} \times 0) = l_{31}l_{21} + l_{32}l_{22} + 0 = l_{31}l_{21} + l_{32}l_{22}$$

Element $R_{33}$:

$$= (l_{31} \times l_{31}) + (l_{32} \times l_{32}) + (l_{33} \times l_{33}) = l_{31}^2 + l_{32}^2 + l_{33}^2$$

So:

$$LL^T = \begin{pmatrix} l_{11}^2 & l_{11}l_{21} & l_{11}l_{31} \\ l_{21}l_{11} & l_{21}^2 + l_{22}^2 & l_{21}l_{31} + l_{22}l_{32} \\ l_{31}l_{11} & l_{31}l_{21} + l_{32}l_{22} & l_{31}^2 + l_{32}^2 + l_{33}^2 \end{pmatrix} = \begin{pmatrix} 4 & 2 & 2 \\ 2 & 2 & 0 \\ 2 & 0 & 2 \end{pmatrix}$$

**Step 3: Next we solve for elements of $L$**

From $R_{11}$:

$$l_{11}^2 = 4 \implies l_{11} = \sqrt{4} = 2$$

From $R_{12}$:

$$l_{11}l_{21} = 2 \implies 2 \times l_{21} = 2 \implies l_{21} = \frac{2}{2} = 1$$

From $R_{13}$:

$$l_{11}l_{31} = 2 \implies 2 \times l_{31} = 2 \implies l_{31} = \frac{2}{2} = 1$$

From $R_{21}$:

$$l_{21}l_{11} = 2 \implies 1 \times 2 = 2 \quad \text{(already satisfied)}$$

From $R_{22}$:

$$l_{21}^2 + l_{22}^2 = 2 \implies (1)^2 + l_{22}^2 = 2 \implies 1 + l_{22}^2 = 2 \implies l_{22}^2 = 1 \implies l_{22} = 1$$

From $R_{23}$:

$$l_{21}l_{31} + l_{22}l_{32} = 0 \implies (1)(1) + (1)l_{32} = 0 \implies 1 + l_{32} = 0 \implies l_{32} = -1$$

From $R_{31}$:

$$l_{31}l_{11} = 2 \implies 1 \times 2 = 2 \quad \text{(already satisfied)}$$

From $R_{32}$:

$$l_{31}l_{21} + l_{32}l_{22} = 0 \implies (1)(1) + (-1)(1) = 1 - 1 = 0 \quad \text{(already satisfied)}$$

From $R_{33}$:

$$l_{31}^2+l_{32}^2+l_{33}^2 = 2 \implies (1)^2+(-1)^2+l_{33}^2 = 2 \implies 1+1+l_{33}^2 = 2 \implies 2+l_{33}^2 = 2 \implies l_{33}^2 = 0$$

**Step 4: Final $L$ matrix and conclusion**

Since $l_{33}^2 = 0$, we have:

$$L = \begin{pmatrix} 2 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & -1 & 0 \end{pmatrix}$$

The Cholesky decomposition fails because we get $l_{33} = 0$, making $L$ singular. Therefore, $C$ is **not positive definite**.

$$\boxed{\text{C is not positive definite; Cholesky yields singular L}}$$

## (d) Explain why a positive definite matrix must have all positive eigenvalues, and relate this to the determinant being positive.

A matrix is positive definite if $x^T C x > 0$ for all nonzero $x$. In spectral decomposition:

$$C = Q\Lambda Q^T,$$

where $\Lambda$ contains eigenvalues. If any eigenvalue $\leq 0$, the quadratic form could be zero or negative. Thus, all eigenvalues must be positive [1]. The determinant is the product of eigenvalues. If all eigenvalues are positive, the determinant is positive.

# Final Answers for Question 1

- $\det(A) = 3,\ \mathrm{tr}(A) = 6,\ A$ is invertible.
- Eigenvalues of $B$: $\dfrac{5 \pm \sqrt{5}}{2}$, sum equals trace, product equals determinant.
- $C$ is not positive definite (Cholesky decomposition fails).
- Positive definite $\Rightarrow$ all eigenvalues positive $\Rightarrow$ determinant positive.

# Question 2: Defective Matrices and Generalized Eigenvectors

We are given:

$$J = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}.$$

## (a) Find eigenvalues and eigenvectors of $J$. What are their algebraic and geometric multiplicities?

**Step 1: Compute characteristic polynomial**

$$\det(J - \lambda I) = \begin{bmatrix} 2-\lambda & 1 \\ 0 & 2-\lambda \end{bmatrix}.$$

Compute determinant:

$$(2-\lambda)(2-\lambda) - (0)(1) = (2-\lambda)^2.$$

So characteristic equation:

$$(2-\lambda)^2 = 0.$$

**Step 2: Solve for eigenvalues**

$$2 - \lambda = 0 \implies \lambda = 2.$$

Thus, the only eigenvalue is:

$$\lambda = 2.$$

**Step 3: Algebraic multiplicity** The factor $(2-\lambda)^2$ shows multiplicity $= 2$:

$$\text{Algebraic multiplicity of } \lambda = 2 \text{ is } 2.$$

**Step 4: To find the eigenvectors**, we can solve:

$$(J - 2I)x = 0.$$

Compute $J - 2I$:

$$J - 2I = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix} - \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

So the system:

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Compute:

$$0 \cdot x_1 + 1 \cdot x_2 = 0 \implies x_2 = 0.$$

No condition on $x_1$, so:

$$x_1 \text{ is free.}$$

Eigenvector:

$$v = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{ (up to scalar multiples).}$$

**Step 5: Geometric multiplicity** Number of independent eigenvectors $= 1$:

$$\text{Geometric multiplicity of } \lambda = 2 \text{ is } 1.$$

## (b) Explain why $J$ is defective and cannot be diagonalized

A matrix is diagonalizable if it has a full set of linearly independent eigenvectors. Recall that:

$$\text{Algebraic multiplicity} = 2, \quad \text{Geometric multiplicity} = 1.$$

Since the geometric multiplicity is less than the algebraic multiplicity, then $J$ does not have enough eigenvectors to form a basis.

$$\therefore\ J \text{ is } \textbf{defective} \text{ and cannot be diagonalized.}$$

## (c) In the context of machine learning, why might we encounter nearly-defective matrices, and what numerical issues can this cause?

In machine learning, **nearly-defective matrices** often arise when a matrix has eigenvalues that are very close together and eigenvectors that are nearly linearly dependent, making the geometric multiplicity almost less than the algebraic multiplicity [3].

**Where do they occur?**

- **Covariance matrices in PCA:** Highly correlated features or multicollinearity cause eigenvalues to cluster near zero, making the matrix ill-conditioned [2].
- **Hessian matrices in optimization:** Near saddle points or regions with highly varying curvature, Hessians become nearly defective [3].
- **Neural network weight matrices:** During training, especially with vanishing or exploding gradients, weight matrices can exhibit near-defective behavior [4].

**Why is this a problem?**

- Small perturbations in input or rounding errors can cause large changes in eigenvectors [4]. Also, algorithms relying on eigen-decomposition (e.g., PCA, optimization) become unstable and unreliable.
- Condition number becomes extremely large, amplifying floating-point errors.

**Consequences:**

- Incorrect solutions and failed convergence in iterative methods like gradient descent.
- Poor generalization and unpredictable model behavior.
- Models may become unsuitable for deployment [2].

**Specific numerical issues include:**

- **Ill-conditioning:** Small input changes cause large output variations [4].
- **Slow convergence:** Iterative algorithms suffer due to poor spectral properties.
- **Instability:** Matrix inversions and decompositions become unreliable.
- **Sensitivity to initialization:** Different random seeds produce dramatically different results.

# Question 3: Matrix Decompositions

We are given:

$$M = \begin{pmatrix} 5 & 2 \\ 2 & 8 \end{pmatrix}.$$

## (a) Perform LU decomposition on $M$ to find lower triangular $L$ and upper triangular $U$ such that $M = LU$.

**Solution:**

We want $M = LU$ where $L$ is lower triangular with 1's on diagonal, $U$ is upper triangular.

Let:

$$L = \begin{pmatrix} 1 & 0 \\ l_{21} & 1 \end{pmatrix}, \quad U = \begin{pmatrix} u_{11} & u_{12} \\ 0 & u_{22} \end{pmatrix}$$

**Step 1: Multiply $L$ and $U$**

$$LU = \begin{pmatrix} 1 & 0 \\ l_{21} & 1 \end{pmatrix} \begin{pmatrix} u_{11} & u_{12} \\ 0 & u_{22} \end{pmatrix}$$

Compute each element:

Element $R_{11}$:

$$= (1 \times u_{11}) + (0 \times 0) = u_{11} + 0 = u_{11}$$

Element $R_{12}$:

$$= (1 \times u_{12}) + (0 \times u_{22}) = u_{12} + 0 = u_{12}$$

Element $R_{21}$:

$$= (l_{21} \times u_{11}) + (1 \times 0) = l_{21}u_{11} + 0 = l_{21}u_{11}$$

Element $R_{22}$:

$$= (l_{21} \times u_{12}) + (1 \times u_{22}) = l_{21}u_{12} + u_{22}$$

So:

$$LU = \begin{pmatrix} u_{11} & u_{12} \\ l_{21}u_{11} & l_{21}u_{12} + u_{22} \end{pmatrix} = \begin{pmatrix} 5 & 2 \\ 2 & 8 \end{pmatrix}$$

**Step 2: Solve for the elements**

From the equality $LU = M$, we equate corresponding elements:
From $R_{11}$:

$$u_{11} = 5$$

From $R_{12}$:

$$u_{12} = 2$$

From $R_{21}$:

$$l_{21}u_{11} = 2$$

Substitute $u_{11} = 5$:

$$l_{21} \times 5 = 2$$

$$l_{21} = \frac{2}{5}$$

From $R_{22}$:

$$l_{21}u_{12} + u_{22} = 8$$

Substitute $l_{21} = \frac{2}{5}$ and $u_{12} = 2$:

$$\left(\frac{2}{5}\right) \times 2 + u_{22} = 8$$

$$\frac{4}{5} + u_{22} = 8$$

$$u_{22} = 8 - \frac{4}{5}$$

Convert to common denominator:

$$8 = \frac{40}{5}, \quad \frac{4}{5} = \frac{4}{5}$$

$$u_{22} = \frac{40}{5} - \frac{4}{5} = \frac{36}{5}$$

**Step 3: Final $L$ and $U$ matrices**

$$L = \begin{pmatrix} 1 & 0 \\ \frac{2}{5} & 1 \end{pmatrix}, \quad U = \begin{pmatrix} 5 & 2 \\ 0 & \frac{36}{5} \end{pmatrix}$$

**Step 4: Verification**
Multiply $L$ and $U$ to verify:

$$LU = \begin{pmatrix} 1 & 0 \\ \frac{2}{5} & 1 \end{pmatrix} \begin{pmatrix} 5 & 2 \\ 0 & \frac{36}{5} \end{pmatrix}$$

Element $R_{11}$:

$$= (1 \times 5) + (0 \times 0) = 5 + 0 = 5$$

Element $R_{12}$:

$$= (1 \times 2) + (0 \times \frac{36}{5}) = 2 + 0 = 2$$

Element $R_{21}$:

$$= \left(\frac{2}{5} \times 5\right) + (1 \times 0) = 2 + 0 = 2$$

Element $R_{22}$:

$$= \left(\frac{2}{5} \times 2\right) + \left(1 \times \frac{36}{5}\right) = \frac{4}{5} + \frac{36}{5} = \frac{40}{5} = 8$$

$$LU = \begin{pmatrix} 5 & 2 \\ 2 & 8 \end{pmatrix} = M$$

$$\boxed{L = \begin{pmatrix} 1 & 0 \\ \frac{2}{5} & 1 \end{pmatrix}, \quad U = \begin{pmatrix} 5 & 2 \\ 0 & \frac{36}{5} \end{pmatrix}}$$

## (b) Perform QR decomposition on $M$ using the Gram-Schmidt process.

**Solution:**

Let columns of $M$ be:

$$a_1 = \begin{pmatrix} 5 \\ 2 \end{pmatrix}, \quad a_2 = \begin{pmatrix} 2 \\ 8 \end{pmatrix}$$

**Step 1: First basis vector**

$$u_1 = a_1 = \begin{pmatrix} 5 \\ 2 \end{pmatrix}$$

Compute norm of $u_1$:

$$\|u_1\| = \sqrt{5^2 + 2^2} = \sqrt{25 + 4} = \sqrt{29}$$

Normalize to get $q_1$:

$$q_1 = \frac{u_1}{\|u_1\|} = \frac{1}{\sqrt{29}} \begin{pmatrix} 5 \\ 2 \end{pmatrix} = \begin{pmatrix} \frac{5}{\sqrt{29}} \\ \frac{2}{\sqrt{29}} \end{pmatrix}$$

**Step 2: Second basis vector**

$$u_2 = a_2 - (a_2 \cdot q_1)q_1$$

First compute $a_2 \cdot q_1$:

$$a_2 \cdot q_1 = \begin{pmatrix} 2 \\ 8 \end{pmatrix} \cdot \begin{pmatrix} \frac{5}{\sqrt{29}} \\ \frac{2}{\sqrt{29}} \end{pmatrix} = 2 \times \frac{5}{\sqrt{29}} + 8 \times \frac{2}{\sqrt{29}} = \frac{10}{\sqrt{29}} + \frac{16}{\sqrt{29}} = \frac{26}{\sqrt{29}}$$

Now compute $(a_2 \cdot q_1)q_1$:

$$(a_2 \cdot q_1)q_1 = \frac{26}{\sqrt{29}} \times \begin{pmatrix} \frac{5}{\sqrt{29}} \\ \frac{2}{\sqrt{29}} \end{pmatrix} = \frac{26}{29} \begin{pmatrix} 5 \\ 2 \end{pmatrix} = \begin{pmatrix} \frac{130}{29} \\ \frac{52}{29} \end{pmatrix}$$

Now compute $u_2$:

$$u_2 = \begin{pmatrix} 2 \\ 8 \end{pmatrix} - \begin{pmatrix} \frac{130}{29} \\ \frac{52}{29} \end{pmatrix} = \begin{pmatrix} 2 - \frac{130}{29} \\ 8 - \frac{52}{29} \end{pmatrix}$$

Convert to common denominator:

$$2 = \frac{58}{29}, \quad 8 = \frac{232}{29}$$

$$u_2 = \begin{pmatrix} \frac{58}{29} - \frac{130}{29} \\ \frac{232}{29} - \frac{52}{29} \end{pmatrix} = \begin{pmatrix} -\frac{72}{29} \\ \frac{180}{29} \end{pmatrix} = \frac{1}{29} \begin{pmatrix} -72 \\ 180 \end{pmatrix}$$

**Step 3: Normalize $u_2$**

Compute norm of $u_2$:

$$\|u_2\| = \left\| \frac{1}{29} \begin{pmatrix} -72 \\ 180 \end{pmatrix} \right\| = \frac{1}{29} \sqrt{(-72)^2 + 180^2}$$

Compute squares:

$$(-72)^2 = 5184, \quad 180^2 = 32400$$

$$5184 + 32400 = 37584$$
$$\sqrt{37584} = \sqrt{36 \times 1044} = \sqrt{36} \times \sqrt{1044} = 6\sqrt{1044}$$
$$\sqrt{1044} = \sqrt{4 \times 261} = \sqrt{4} \times \sqrt{261} = 2\sqrt{261}$$
$$\sqrt{37584} = 6 \times 2\sqrt{261} = 12\sqrt{261}$$

Therefore:

$$\|u_2\| = \frac{1}{29} \times 12\sqrt{261} = \frac{12\sqrt{261}}{29}$$

Now compute $q_2$:

$$q_2 = \frac{u_2}{\|u_2\|} = \frac{\frac{1}{29}\begin{pmatrix}-72\\180\end{pmatrix}}{\frac{12\sqrt{261}}{29}} = \frac{1}{12\sqrt{261}}\begin{pmatrix}-72\\180\end{pmatrix}$$

Simplify:

$$\frac{-72}{12\sqrt{261}} = -\frac{6}{\sqrt{261}}, \quad \frac{180}{12\sqrt{261}} = \frac{15}{\sqrt{261}}$$
$$q_2 = \begin{pmatrix}-\frac{6}{\sqrt{261}}\\ \frac{15}{\sqrt{261}}\end{pmatrix}$$

**Step 4: Construct Q and R matrices**

**Q matrix:**

$$Q = \begin{pmatrix}\frac{5}{\sqrt{29}} & -\frac{6}{\sqrt{261}}\\ \frac{2}{\sqrt{29}} & \frac{15}{\sqrt{261}}\end{pmatrix}$$

**R matrix (upper triangular with $r_{ij} = a_j \cdot q_i$):**

Compute $r_{11} = a_1 \cdot q_1$:

$$r_{11} = \begin{pmatrix}5\\2\end{pmatrix} \cdot \begin{pmatrix}\frac{5}{\sqrt{29}}\\ \frac{2}{\sqrt{29}}\end{pmatrix} = 5 \times \frac{5}{\sqrt{29}} + 2 \times \frac{2}{\sqrt{29}} = \frac{25}{\sqrt{29}} + \frac{4}{\sqrt{29}} = \frac{29}{\sqrt{29}} = \sqrt{29}$$

Compute $r_{12} = a_2 \cdot q_1$:

$$r_{12} = \begin{pmatrix}2\\8\end{pmatrix} \cdot \begin{pmatrix}\frac{5}{\sqrt{29}}\\ \frac{2}{\sqrt{29}}\end{pmatrix} = 2 \times \frac{5}{\sqrt{29}} + 8 \times \frac{2}{\sqrt{29}} = \frac{10}{\sqrt{29}} + \frac{16}{\sqrt{29}} = \frac{26}{\sqrt{29}}$$

Compute $r_{22} = a_2 \cdot q_2$:

$$r_{22} = \begin{pmatrix} 2 \\ 8 \end{pmatrix} \cdot \begin{pmatrix} -\frac{6}{\sqrt{261}} \\ \frac{15}{\sqrt{261}} \end{pmatrix} = 2 \times \left(-\frac{6}{\sqrt{261}}\right) + 8 \times \frac{15}{\sqrt{261}} = -\frac{12}{\sqrt{261}} + \frac{120}{\sqrt{261}} = \frac{108}{\sqrt{261}}$$

Simplify $r_{22}$:

$$\frac{108}{\sqrt{261}} = \frac{108\sqrt{261}}{261} = \frac{36\sqrt{261}}{87} = \frac{12\sqrt{261}}{29}$$

**Final R matrix:**

$$R = \begin{pmatrix} \sqrt{29} & \frac{26}{\sqrt{29}} \\ 0 & \frac{12\sqrt{261}}{29} \end{pmatrix}$$

**Verification:** $QR = M$

$$QR = \begin{pmatrix} \frac{5}{\sqrt{29}} & -\frac{6}{\sqrt{261}} \\ \frac{2}{\sqrt{29}} & \frac{15}{\sqrt{261}} \end{pmatrix} \begin{pmatrix} \sqrt{29} & \frac{26}{\sqrt{29}} \\ 0 & \frac{12\sqrt{261}}{29} \end{pmatrix}$$

Element $R_{11}$:

$$= \frac{5}{\sqrt{29}} \times \sqrt{29} + \left(-\frac{6}{\sqrt{261}}\right) \times 0 = 5 + 0 = 5$$

Element $R_{12}$:

$$= \frac{5}{\sqrt{29}} \times \frac{26}{\sqrt{29}} + \left(-\frac{6}{\sqrt{261}}\right) \times \frac{12\sqrt{261}}{29} = \frac{130}{29} - \frac{72}{29} = \frac{58}{29} = 2$$

Element $R_{21}$:

$$= \frac{2}{\sqrt{29}} \times \sqrt{29} + \frac{15}{\sqrt{261}} \times 0 = 2 + 0 = 2$$

Element $R_{22}$:

$$= \frac{2}{\sqrt{29}} \times \frac{26}{\sqrt{29}} + \frac{15}{\sqrt{261}} \times \frac{12\sqrt{261}}{29} = \frac{52}{29} + \frac{180}{29} = \frac{232}{29} = 8$$

$$QR = \begin{pmatrix} 5 & 2 \\ 2 & 8 \end{pmatrix} = M$$

$$\boxed{Q = \begin{pmatrix} \frac{5}{\sqrt{29}} & -\frac{6}{\sqrt{261}} \\ \frac{2}{\sqrt{29}} & \frac{15}{\sqrt{261}} \end{pmatrix}, \quad R = \begin{pmatrix} \sqrt{29} & \frac{26}{\sqrt{29}} \\ 0 & \frac{12\sqrt{261}}{29} \end{pmatrix}}$$

## (c) Compute eigendecomposition and verify $PDP^{-1} = M$

Compute characteristic polynomial:

$$\det(M - \lambda I) = \det \begin{bmatrix} 5-\lambda & 2 \\ 2 & 8-\lambda \end{bmatrix} = (5-\lambda)(8-\lambda) - (2)(2).$$

Expand $(5-\lambda)(8-\lambda)$:

$$\begin{aligned} &= 5 \times 8 + 5 \times (-\lambda) + (-\lambda) \times 8 + (-\lambda) \times (-\lambda) \\ &= 40 - 5\lambda - 8\lambda + \lambda^2 \\ &= \lambda^2 - 13\lambda + 40. \end{aligned}$$

Subtract 4:

$$\lambda^2 - 13\lambda + 40 - 4 = \lambda^2 - 13\lambda + 36.$$

So characteristic equation:

$$\lambda^2 - 13\lambda + 36 = 0.$$

Solve using quadratic formula:

$$\begin{aligned} \lambda &= \frac{-(-13) \pm \sqrt{(-13)^2 - 4(1)(36)}}{2(1)} \\ &= \frac{13 \pm \sqrt{169 - 144}}{2} \\ &= \frac{13 \pm \sqrt{25}}{2} \\ &= \frac{13 \pm 5}{2}. \end{aligned}$$

Calculate both eigenvalues:

$$\lambda_1 = \frac{13+5}{2} = \frac{18}{2} = 9$$

$$\lambda_2 = \frac{13-5}{2} = \frac{8}{2} = 4.$$

**Step 1: Find eigenvectors and normalize them**

For $\lambda_1 = 9$:

$$(M - 9I) = \begin{pmatrix} 5-9 & 2 \\ 2 & 8-9 \end{pmatrix} = \begin{pmatrix} -4 & 2 \\ 2 & -1 \end{pmatrix}.$$

Solve $(M - 9I)v = 0$:

$$\begin{pmatrix} -4 & 2 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

First equation: $-4x_1 + 2x_2 = 0$

$$2x_2 = 4x_1$$

$$x_2 = 2x_1.$$

Choose $x_1 = 1$, then $x_2 = 2$:

$$v_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

Normalize $v_1$:

$$\|v_1\| = \sqrt{1^2 + 2^2} = \sqrt{1+4} = \sqrt{5}.$$

$$p_1 = \frac{v_1}{\|v_1\|} = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{pmatrix}.$$

For $\lambda_2 = 4$:

$$(M - 4I) = \begin{pmatrix} 5-4 & 2 \\ 2 & 8-4 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}.$$

Solve $(M - 4I)v = 0$:

$$\begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

First equation: $1x_1 + 2x_2 = 0$

$$x_1 = -2x_2.$$

Choose $x_2 = 1$, then $x_1 = -2$:

$$v_2 = \begin{pmatrix} -2 \\ 1 \end{pmatrix}.$$

Normalize $v_2$:

$$\|v_2\| = \sqrt{(-2)^2 + 1^2} = \sqrt{4+1} = \sqrt{5}.$$

$$p_2 = \frac{v_2}{\|v_2\|} = \frac{1}{\sqrt{5}} \begin{pmatrix} -2 \\ 1 \end{pmatrix} = \begin{pmatrix} -\frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix}.$$

**Step 2: Construct normalized matrices**

$$P = \begin{pmatrix} \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix}, \; D = \begin{pmatrix} 9 & 0 \\ 0 & 4 \end{pmatrix}.$$

**Step 3: Verification: Compute $PDP^{-1}$**
Since $P$ has orthonormal columns, $P^{-1} = P^T$:

$$P^{-1} = P^T = \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix}.$$

Compute $PD$:

$$PD = \begin{pmatrix} \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} 9 & 0 \\ 0 & 4 \end{pmatrix}.$$

Element $R_{11}$ of $PD$:

$$= \left(\frac{1}{\sqrt{5}}\right)(9) + \left(-\frac{2}{\sqrt{5}}\right)(0) = \frac{9}{\sqrt{5}} + 0 = \frac{9}{\sqrt{5}}.$$

Element $R_{12}$ of $PD$:

$$= \left(\frac{1}{\sqrt{5}}\right)(0) + \left(-\frac{2}{\sqrt{5}}\right)(4) = 0 + \left(-\frac{8}{\sqrt{5}}\right) = -\frac{8}{\sqrt{5}}.$$

Element $R_{21}$ of $PD$:

$$= \left(\frac{2}{\sqrt{5}}\right)(9) + \left(\frac{1}{\sqrt{5}}\right)(0) = \frac{18}{\sqrt{5}} + 0 = \frac{18}{\sqrt{5}}.$$

Element $R_{22}$ of $PD$:

$$= \left(\frac{2}{\sqrt{5}}\right)(0) + \left(\frac{1}{\sqrt{5}}\right)(4) = 0 + \frac{4}{\sqrt{5}} = \frac{4}{\sqrt{5}}.$$

So:

$$PD = \begin{pmatrix} \frac{9}{\sqrt{5}} & -\frac{8}{\sqrt{5}} \\ \frac{18}{\sqrt{5}} & \frac{4}{\sqrt{5}} \end{pmatrix}.$$

Now compute $(PD)P^{-1}$:

$$(PD)P^{-1} = \begin{pmatrix} \frac{9}{\sqrt{5}} & -\frac{8}{\sqrt{5}} \\ \frac{18}{\sqrt{5}} & \frac{4}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix}.$$

Compute element $R_{11}$:

$$\begin{aligned} &= \left(\frac{9}{\sqrt{5}}\right)\left(\frac{1}{\sqrt{5}}\right) + \left(-\frac{8}{\sqrt{5}}\right)\left(-\frac{2}{\sqrt{5}}\right) \\ &= \frac{9}{\sqrt{5}\cdot\sqrt{5}} + \frac{16}{\sqrt{5}\cdot\sqrt{5}} \\ &= \frac{9}{5} + \frac{16}{5} = \frac{25}{5} = 5. \end{aligned}$$

Compute element $R_{12}$:

$$\begin{aligned} &= \left(\frac{9}{\sqrt{5}}\right)\left(\frac{2}{\sqrt{5}}\right) + \left(-\frac{8}{\sqrt{5}}\right)\left(\frac{1}{\sqrt{5}}\right) \\ &= \frac{18}{\sqrt{5}\cdot\sqrt{5}} + \left(-\frac{8}{\sqrt{5}\cdot\sqrt{5}}\right) \\ &= \frac{18}{5} - \frac{8}{5} = \frac{10}{5} = 2. \end{aligned}$$

Compute element $R_{21}$:

$$\begin{aligned} &= \left(\frac{18}{\sqrt{5}}\right)\left(\frac{1}{\sqrt{5}}\right) + \left(\frac{4}{\sqrt{5}}\right)\left(-\frac{2}{\sqrt{5}}\right) \\ &= \frac{18}{\sqrt{5}\cdot\sqrt{5}} + \left(-\frac{8}{\sqrt{5}\cdot\sqrt{5}}\right) \\ &= \frac{18}{5} - \frac{8}{5} = \frac{10}{5} = 2. \end{aligned}$$

Compute element $R_{22}$:

$$= \left(\frac{18}{\sqrt{5}}\right)\left(\frac{2}{\sqrt{5}}\right) + \left(\frac{4}{\sqrt{5}}\right)\left(\frac{1}{\sqrt{5}}\right)$$

$$= \frac{36}{\sqrt{5} \cdot \sqrt{5}} + \frac{4}{\sqrt{5} \cdot \sqrt{5}}$$
$$= \frac{36}{5} + \frac{4}{5} = \frac{40}{5} = 8.$$

Therefore:

$$PDP^{-1} = \begin{pmatrix} 5 & 2 \\ 2 & 8 \end{pmatrix} = M.$$

$$\boxed{P = \begin{pmatrix} \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix}, \ D = \begin{pmatrix} 9 & 0 \\ 0 & 4 \end{pmatrix}, \ PDP^{-1} = M}$$

Verified: $PDP^{-1} = M$ with normalized eigenvectors.

### (d) Compare decompositions

- **LU**: Reveals triangular structure, efficient for solving $Mx = b$ repeatedly.
- **QR**: Reveals orthogonal basis, useful for least squares problems.
- **Eigen**: Reveals spectral properties, useful for diagonalization and understanding variance.
- For solving $Mx = b$, LU decomposition is most efficient.

## Final Answers for Question 3

- LU: $L = \begin{pmatrix} 1 & 0 \\ \frac{2}{5} & 1 \end{pmatrix}, \quad U = \begin{pmatrix} 5 & 2 \\ 0 & \frac{36}{5} \end{pmatrix}$
- QR: $\boxed{Q = \begin{pmatrix} \frac{5}{\sqrt{29}} & -\frac{6}{\sqrt{261}} \\ \frac{2}{\sqrt{29}} & \frac{15}{\sqrt{261}} \end{pmatrix}, \quad R = \begin{pmatrix} \sqrt{29} & \frac{26}{\sqrt{29}} \\ 0 & \frac{12\sqrt{261}}{29} \end{pmatrix}}.$
- Eigen: $P = \begin{pmatrix} \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix}, \ D = \begin{pmatrix} 9 & 0 \\ 0 & 4 \end{pmatrix}$, and $PDP^{-1} = M$ verified.
- LU is most efficient for solving $Mx = b$.

# Question 4: Singular Value Decomposition and PCA [20 Marks]

Consider the data matrix:

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

## (a) Compute the singular value decomposition (SVD) of $X$ by hand. Show all steps, including finding eigenvalues/eigenvectors of $X^\top X$ and $XX^\top$.

**Solution:**

**Step 1: Compute $X^T X$**

First, find $X^T$ (transpose of X):

$$X^T = \begin{pmatrix} 1 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix}$$

Now compute $X^T X$:

$$X^T X = \begin{pmatrix} 1 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix}$$

Element $R_{11}$:

$$= (1 \times 1) + (1 \times 1) + (0 \times 0) = 1 + 1 + 0 = 2$$

Element $R_{12}$:

$$= (1 \times 2) + (1 \times 0) + (0 \times 1) = 2 + 0 + 0 = 2$$

Element $R_{21}$:

$$= (2 \times 1) + (0 \times 1) + (1 \times 0) = 2 + 0 + 0 = 2$$

Element $R_{22}$:

$$= (2 \times 2) + (0 \times 0) + (1 \times 1) = 4 + 0 + 1 = 5$$

So:

$$X^T X = \begin{pmatrix} 2 & 2 \\ 2 & 5 \end{pmatrix}$$

**Step 2: Compute $XX^T$**

$$XX^T = \begin{pmatrix} 1 & 2 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix} = \begin{pmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{pmatrix}$$

Element $R_{11}$:

$$= (1 \times 1) + (2 \times 2) = 1 + 4 = 5$$

Element $R_{12}$:

$$= (1 \times 1) + (2 \times 0) = 1 + 0 = 1$$

Element $R_{13}$:

$$= (1 \times 0) + (2 \times 1) = 0 + 2 = 2$$

Element $R_{21}$:

$$= (1 \times 1) + (0 \times 2) = 1 + 0 = 1$$

Element $R_{22}$:

$$= (1 \times 1) + (0 \times 0) = 1 + 0 = 1$$

Element $R_{23}$:

$$= (1 \times 0) + (0 \times 1) = 0 + 0 = 0$$

Element $R_{31}$:

$$= (0 \times 1) + (1 \times 2) = 0 + 2 = 2$$

Element $R_{32}$:

$$= (0 \times 1) + (1 \times 0) = 0 + 0 = 0$$

Element $R_{33}$:

$$= (0 \times 0) + (1 \times 1) = 0 + 1 = 1$$

So:

$$XX^T = \begin{pmatrix} 5 & 1 & 2 \\ 1 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix}$$

**Step 3: Find eigenvalues of $X^TX$**

First, we solve: $\det(X^TX - \lambda I) = 0$

$$X^TX - \lambda I = \begin{pmatrix} 2-\lambda & 2 \\ 2 & 5-\lambda \end{pmatrix}$$

Determinant calculation:

$$\begin{aligned} \det &= (2-\lambda)(5-\lambda) - (2)(2) \\ &= (10 - 2\lambda - 5\lambda + \lambda^2) - 4 \\ &= 10 - 7\lambda + \lambda^2 - 4 \\ &= \lambda^2 - 7\lambda + 6 \end{aligned}$$

Solve quadratic equation:

$$\lambda^2 - 7\lambda + 6 = 0$$

$$\lambda = \frac{7 \pm \sqrt{(-7)^2 - 4(1)(6)}}{2(1)} = \frac{7 \pm \sqrt{49-24}}{2} = \frac{7 \pm \sqrt{25}}{2} = \frac{7 \pm 5}{2}$$

$$\lambda_1 = \frac{7+5}{2} = \frac{12}{2} = 6, \quad \lambda_2 = \frac{7-5}{2} = \frac{2}{2} = 1$$

**Step 4: Find eigenvectors of $X^TX$**

For $\lambda_1 = 6$:

$$(X^TX - 6I)v = \begin{pmatrix} 2-6 & 2 \\ 2 & 5-6 \end{pmatrix} v = \begin{pmatrix} -4 & 2 \\ 2 & -1 \end{pmatrix} v = 0$$

Write as system of equations:

$$-4v_1 + 2v_2 = 0 \quad (1)$$

$$2v_1 - 1v_2 = 0 \quad (2)$$

From equation (1): $-4v_1 + 2v_2 = 0 \Rightarrow 2v_2 = 4v_1 \Rightarrow v_2 = 2v_1$
Choose $v_1 = 1$, then $v_2 = 2$
So eigenvector: $v_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$

For $\lambda_2 = 1$:

$$(X^T X - I)v = \begin{pmatrix} 2-1 & 2 \\ 2 & 5-1 \end{pmatrix} v = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} v = 0$$

Write as system of equations:

$$1v_1 + 2v_2 = 0 \quad (1)$$

$$2v_1 + 4v_2 = 0 \quad (2)$$

From equation (1): $v_1 + 2v_2 = 0 \Rightarrow v_1 = -2v_2$
Choose $v_2 = 1$, then $v_1 = -2$
So eigenvector: $v_2 = \begin{pmatrix} -2 \\ 1 \end{pmatrix}$

**Step 5: Normalize eigenvectors to get V matrix**

Normalize $v_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$:

$$\|v_1\| = \sqrt{1^2 + 2^2} = \sqrt{1+4} = \sqrt{5}$$

$$\hat{v}_1 = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{pmatrix}$$

Normalize $v_2 = \begin{pmatrix} -2 \\ 1 \end{pmatrix}$:

$$\|v_2\| = \sqrt{(-2)^2 + 1^2} = \sqrt{4+1} = \sqrt{5}$$

$$\hat{v}_2 = \frac{1}{\sqrt{5}} \begin{pmatrix} -2 \\ 1 \end{pmatrix} = \begin{pmatrix} -\frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix}$$

So V matrix:

$$V = \begin{pmatrix} \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix}$$

**Step 6: Construct $\Sigma$ matrix from singular values**

Singular values: $\sigma_i = \sqrt{\lambda_i}$

$$\sigma_1 = \sqrt{6}, \quad \sigma_2 = \sqrt{1} = 1$$

$\Sigma$ matrix (3×2):

$$\Sigma = \begin{pmatrix} \sqrt{6} & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$$

**Step 7: Compute U matrix columns**

Compute $u_1 = \frac{1}{\sigma_1} X v_1$:
First compute $X v_1$:

$$X v_1 = \begin{pmatrix} 1 & 2 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{pmatrix} = \begin{pmatrix} 1 \times \frac{1}{\sqrt{5}} + 2 \times \frac{2}{\sqrt{5}} \\ 1 \times \frac{1}{\sqrt{5}} + 0 \times \frac{2}{\sqrt{5}} \\ 0 \times \frac{1}{\sqrt{5}} + 1 \times \frac{2}{\sqrt{5}} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{5}} + \frac{4}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} + 0 \\ 0 + \frac{2}{\sqrt{5}} \end{pmatrix} = \begin{pmatrix} \frac{5}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{pmatrix}$$

Now multiply by $\frac{1}{\sigma_1} = \frac{1}{\sqrt{6}}$:

$$u_1 = \frac{1}{\sqrt{6}} \begin{pmatrix} \frac{5}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{pmatrix} = \begin{pmatrix} \frac{5}{\sqrt{30}} \\ \frac{1}{\sqrt{30}} \\ \frac{2}{\sqrt{30}} \end{pmatrix}$$

Compute $u_2 = \frac{1}{\sigma_2} X v_2$:
First compute $X v_2$:

$$X v_2 = \begin{pmatrix} 1 & 2 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} -\frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix} = \begin{pmatrix} 1 \times (-\frac{2}{\sqrt{5}}) + 2 \times \frac{1}{\sqrt{5}} \\ 1 \times (-\frac{2}{\sqrt{5}}) + 0 \times \frac{1}{\sqrt{5}} \\ 0 \times (-\frac{2}{\sqrt{5}}) + 1 \times \frac{1}{\sqrt{5}} \end{pmatrix} = \begin{pmatrix} -\frac{2}{\sqrt{5}} + \frac{2}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} + 0 \\ 0 + \frac{1}{\sqrt{5}} \end{pmatrix} = \begin{pmatrix} 0 \\ -\frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix}$$

Now multiply by $\frac{1}{\sigma_2} = \frac{1}{1} = 1$:

$$u_2 = \begin{pmatrix} 0 \\ -\frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix}$$

Find $u_3$ orthogonal to $u_1$ and $u_2$ using cross product:

Let $a = u_1 = \begin{pmatrix} \frac{5}{\sqrt{30}} \\ \frac{1}{\sqrt{30}} \\ \frac{2}{\sqrt{30}} \end{pmatrix}$, $b = u_2 = \begin{pmatrix} 0 \\ -\frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix}$

Cross product:

$$a \times b = \begin{bmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{5}{\sqrt{30}} & \frac{1}{\sqrt{30}} & \frac{2}{\sqrt{30}} \\ 0 & -\frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix}$$

Calculate determinant:

$$= \mathbf{i}\left(\frac{1}{\sqrt{30}} \times \frac{1}{\sqrt{5}} - \frac{2}{\sqrt{30}} \times (-\frac{2}{\sqrt{5}})\right) - \mathbf{j}\left(\frac{5}{\sqrt{30}} \times \frac{1}{\sqrt{5}} - \frac{2}{\sqrt{30}} \times 0\right) + \mathbf{k}\left(\frac{5}{\sqrt{30}} \times (-\frac{2}{\sqrt{5}}) - \right.$$

$$= \mathbf{i}\left(\frac{1}{\sqrt{150}} + \frac{4}{\sqrt{150}}\right) - \mathbf{j}\left(\frac{5}{\sqrt{150}} - 0\right) + \mathbf{k}\left(-\frac{10}{\sqrt{150}} - 0\right)$$

$$= \mathbf{i}\left(\frac{5}{\sqrt{150}}\right) - \mathbf{j}\left(\frac{5}{\sqrt{150}}\right) + \mathbf{k}\left(-\frac{10}{\sqrt{150}}\right) = \frac{1}{\sqrt{150}}\begin{pmatrix} 5 \\ -5 \\ -10 \end{pmatrix}$$

Normalize: $\|a \times b\| = \frac{1}{\sqrt{150}}\sqrt{25 + 25 + 100} = \frac{1}{\sqrt{150}}\sqrt{150} = 1$
So:

$$u_3 = \frac{1}{\sqrt{150}}\begin{pmatrix} 5 \\ -5 \\ -10 \end{pmatrix} = \frac{1}{\sqrt{6}}\begin{pmatrix} 1 \\ -1 \\ -2 \end{pmatrix}$$

Final U matrix:

$$U = \begin{pmatrix} \frac{5}{\sqrt{30}} & 0 & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{30}} & -\frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{30}} & \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{6}} \end{pmatrix}$$

Final SVD: $X = U\Sigma V^T$

**Step 8: Find eigenvalues and eigenvectors of $XX^T$**
We have:

$$XX^T = \begin{pmatrix} 5 & 1 & 2 \\ 1 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix}$$

Find eigenvalues by solving $\det(XX^T - \lambda I) = 0$:

$$XX^T - \lambda I = \begin{pmatrix} 5-\lambda & 1 & 2 \\ 1 & 1-\lambda & 0 \\ 2 & 0 & 1-\lambda \end{pmatrix}$$

Compute determinant using first row expansion:

$$\det = (5-\lambda)\begin{vmatrix} 1-\lambda & 0 \\ 0 & 1-\lambda \end{vmatrix} - (1)\begin{vmatrix} 1 & 0 \\ 2 & 1-\lambda \end{vmatrix} + (2)\begin{vmatrix} 1 & 1-\lambda \\ 2 & 0 \end{vmatrix}$$

**Compute first 2×2 determinant:**

$$\begin{vmatrix} 1-\lambda & 0 \\ 0 & 1-\lambda \end{vmatrix} = (1-\lambda) \times (1-\lambda) - 0 \times 0$$

$$= (1-\lambda)(1-\lambda) - 0 = (1-\lambda)^2$$

**Compute second 2×2 determinant:**

$$\begin{vmatrix} 1 & 0 \\ 2 & 1-\lambda \end{vmatrix} = 1 \times (1-\lambda) - 0 \times 2$$

$$= (1-\lambda) - 0 = 1-\lambda$$

**Compute third 2×2 determinant:**

$$\begin{vmatrix} 1 & 1-\lambda \\ 2 & 0 \end{vmatrix} = 1 \times 0 - (1-\lambda) \times 2$$

$$= 0 - 2(1-\lambda) = -2(1-\lambda)$$

**Now substitute back into determinant formula:**

$$\det = (5-\lambda)(1-\lambda)^2 - (1)(1-\lambda) + (2)(-2(1-\lambda))$$

$$= (5-\lambda)(1-\lambda)^2 - (1-\lambda) - 4(1-\lambda)$$

**Combine like terms:**

$$-(1-\lambda) - 4(1-\lambda) = -5(1-\lambda)$$

So:

$$\det = (5-\lambda)(1-\lambda)^2 - 5(1-\lambda)$$

**Factor out $(1-\lambda)$:**

$$\det = (1-\lambda)\left[(5-\lambda)(1-\lambda) - 5\right]$$

**Expand inside brackets:**

$$(5-\lambda)(1-\lambda) = 5 \times 1 + 5 \times (-\lambda) + (-\lambda) \times 1 + (-\lambda) \times (-\lambda)$$

$$= 5 - 5\lambda - \lambda + \lambda^2 = 5 - 6\lambda + \lambda^2$$

Now subtract 5:

$$(5 - 6\lambda + \lambda^2) - 5 = -6\lambda + \lambda^2 = \lambda^2 - 6\lambda$$

So:

$$\det = (1 - \lambda)(\lambda^2 - 6\lambda)$$

**Factor $\lambda^2 - 6\lambda$:**

$$\lambda^2 - 6\lambda = \lambda(\lambda - 6)$$

So:

$$\det = (1 - \lambda)\lambda(\lambda - 6)$$

Set determinant to zero:

$$(1 - \lambda)\lambda(\lambda - 6) = 0$$

**Solve for eigenvalues:**

$$1 - \lambda = 0 \Rightarrow \lambda = 1$$

$$\lambda = 0 \Rightarrow \lambda = 0$$

$$\lambda - 6 = 0 \Rightarrow \lambda = 6$$

So eigenvalues:

$$\lambda_1 = 6, \quad \lambda_2 = 1, \quad \lambda_3 = 0$$

**Step 9: Find eigenvectors of $XX^T$**
**For $\lambda_1 = 6$:**

$$XX^T - 6I = \begin{pmatrix} 5-6 & 1 & 2 \\ 1 & 1-6 & 0 \\ 2 & 0 & 1-6 \end{pmatrix} = \begin{pmatrix} -1 & 1 & 2 \\ 1 & -5 & 0 \\ 2 & 0 & -5 \end{pmatrix}$$

Solve $(XX^T - 6I)u = 0$:

$$\begin{pmatrix} -1 & 1 & 2 \\ 1 & -5 & 0 \\ 2 & 0 & -5 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Write as system of equations:

$$\begin{cases} -1u_1 + 1u_2 + 2u_3 = 0 & (1) \\ 1u_1 - 5u_2 + 0u_3 = 0 & (2) \\ 2u_1 + 0u_2 - 5u_3 = 0 & (3) \end{cases}$$

From equation (2):

$$u_1 - 5u_2 = 0 \Rightarrow u_1 = 5u_2$$

From equation (3):

$$2u_1 - 5u_3 = 0 \Rightarrow 5u_3 = 2u_1 \Rightarrow u_3 = \frac{2}{5}u_1$$

Substitute $u_1 = 5u_2$:

$$u_3 = \frac{2}{5}(5u_2) = 2u_2$$

Check equation (1):

$$-1u_1 + 1u_2 + 2u_3 = -1(5u_2) + 1u_2 + 2(2u_2) = -5u_2 + u_2 + 4u_2 = 0u_2 = 0$$

Choose $u_2 = 1$, then:

$$u_1 = 5 \times 1 = 5, \quad u_3 = 2 \times 1 = 2$$

So eigenvector: $u_1 = \begin{pmatrix} 5 \\ 1 \\ 2 \end{pmatrix}$

**For $\lambda_2 = 1$:**

$$XX^T - I = \begin{pmatrix} 5-1 & 1 & 2 \\ 1 & 1-1 & 0 \\ 2 & 0 & 1-1 \end{pmatrix} = \begin{pmatrix} 4 & 1 & 2 \\ 1 & 0 & 0 \\ 2 & 0 & 0 \end{pmatrix}$$

Solve $(XX^T - I)u = 0$:

$$\begin{pmatrix} 4 & 1 & 2 \\ 1 & 0 & 0 \\ 2 & 0 & 0 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Write as system of equations:

$$\begin{cases} 4u_1 + 1u_2 + 2u_3 = 0 & (1) \\ 1u_1 + 0u_2 + 0u_3 = 0 & (2) \\ 2u_1 + 0u_2 + 0u_3 = 0 & (3) \end{cases}$$

From equation (2):

$$u_1 + 0 + 0 = 0 \Rightarrow u_1 = 0$$

From equation (3):

$$2u_1 + 0 + 0 = 0 \Rightarrow 2 \times 0 = 0$$

Now use equation (1) with $u_1 = 0$:

$$4(0) + 1u_2 + 2u_3 = 0 \Rightarrow u_2 + 2u_3 = 0 \Rightarrow u_2 = -2u_3$$

Choose $u_3 = 1$, then:

$$u_2 = -2 \times 1 = -2$$

So eigenvector: $u_2 = \begin{pmatrix} 0 \\ -2 \\ 1 \end{pmatrix}$

**For $\lambda_3 = 0$:**

$$XX^T - 0I = \begin{pmatrix} 5 & 1 & 2 \\ 1 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix}$$

Solve $XX^T u = 0$:

$$\begin{pmatrix} 5 & 1 & 2 \\ 1 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Write as system of equations:

$$\begin{cases} 5u_1 + 1u_2 + 2u_3 = 0 & (1) \\ 1u_1 + 1u_2 + 0u_3 = 0 & (2) \\ 2u_1 + 0u_2 + 1u_3 = 0 & (3) \end{cases}$$

From equation (2):

$$u_1 + u_2 = 0 \Rightarrow u_2 = -u_1$$

From equation (3):

$$2u_1 + u_3 = 0 \Rightarrow u_3 = -2u_1$$

Now use equation (1) to verify:

$$5u_1 + 1(-u_1) + 2(-2u_1) = 5u_1 - u_1 - 4u_1 = 0u_1 = 0$$

Choose $u_1 = 1$, then:

$$u_2 = -1 \times 1 = -1, \quad u_3 = -2 \times 1 = -2$$

So eigenvector: $u_3 = \begin{pmatrix} 1 \\ -1 \\ -2 \end{pmatrix}$

**$X^TX$ and $XX^T$ eigenvectors**

The eigenvalues of $XX^T$ are: $\lambda = 6, 1, 0$
The eigenvalues of $X^TX$ were: $\lambda = 6, 1$
The singular values from $X^TX$ were: $\sigma_1 = \sqrt{6}, \sigma_2 = 1$

**This completes the full SVD with eigenvalues/eigenvectors of both $X^TX$ and $XX^T$.**

## (b) What is the rank of $X$? Write down orthonormal bases for the four fundamental subspaces of $X$ using the SVD components.

**Rank of X:**

- Since the number of nonzero singular values $= 2$ ($\sigma_1 = \sqrt{6} \neq 0$, $\sigma_2 = 1 \neq 0$), therefore:

$$\boxed{\text{The rank}(X) = 2}$$

**Four Fundamental Subspaces:**

1. **Column Space (Range):**

$$\text{span}\left\{ \begin{pmatrix} \frac{5}{\sqrt{30}} \\ \frac{1}{\sqrt{30}} \\ \frac{2}{\sqrt{30}} \end{pmatrix}, \begin{pmatrix} 0 \\ -\frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix} \right\}$$

2. **Row Space:**

$$\text{span}\left\{\begin{pmatrix}\frac{1}{\sqrt{5}}\\ \frac{2}{\sqrt{5}}\end{pmatrix}, \begin{pmatrix}-\frac{2}{\sqrt{5}}\\ \frac{1}{\sqrt{5}}\end{pmatrix}\right\}$$

3. **Left Nullspace:**

$$\text{span}\left\{\begin{pmatrix}\frac{1}{\sqrt{6}}\\ -\frac{1}{\sqrt{6}}\\ -\frac{2}{\sqrt{6}}\end{pmatrix}\right\}$$

**Null space** $\mathcal{N}(X)$: since $\text{rank}(X) = 2$ and $X \in \mathbb{R}^{3\times 2}$, $\dim \mathcal{N}(X) = 2 - 2 = 0$:

$$\boxed{\mathcal{N}(X) = \{\mathbf{0}\} \text{ (trivial; no nonzero basis vector).}}$$

**Left null space** $\mathcal{N}(X^\top) = \text{span}\{\hat{u}_3\}$:

$$\boxed{\mathcal{N}(X^\top) = \text{span}\left\{\begin{pmatrix}\frac{1}{\sqrt{6}}\\ -\frac{1}{\sqrt{6}}\\ -\frac{2}{\sqrt{6}}\end{pmatrix}\right\}} \quad \text{(orthonormal).}$$

## (c) Perform PCA on this dataset by: (i) Centering the data (ii) Computing the covariance matrix (iii) Finding the principal components and proportion of variance explained

### (i) Center the data

Compute column means:

- Column 1 mean: $\frac{1+1+0}{3} = \frac{2}{3}$
- Column 2 mean: $\frac{2+0+1}{3} = \frac{3}{3} = 1$

Center the data:

$$X_c = X - \text{column means} = \begin{pmatrix}1 & 2\\ 1 & 0\\ 0 & 1\end{pmatrix} - \begin{pmatrix}\frac{2}{3} & 1\\ \frac{2}{3} & 1\\ \frac{2}{3} & 1\end{pmatrix} = \begin{pmatrix}1-\frac{2}{3} & 2-1\\ 1-\frac{2}{3} & 0-1\\ 0-\frac{2}{3} & 1-1\end{pmatrix} = \begin{pmatrix}\frac{1}{3} & 1\\ \frac{1}{3} & -1\\ -\frac{2}{3} & 0\end{pmatrix}$$

**(ii) Compute covariance matrix**

First compute $X_c^T X_c$:

$$X_c^T X_c = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & -\frac{2}{3} \\ 1 & -1 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{3} & 1 \\ \frac{1}{3} & -1 \\ -\frac{2}{3} & 0 \end{pmatrix}$$

Element $R_{11}$:

$$= (\frac{1}{3} \times \frac{1}{3}) + (\frac{1}{3} \times \frac{1}{3}) + (-\frac{2}{3} \times -\frac{2}{3}) = \frac{1}{9} + \frac{1}{9} + \frac{4}{9} = \frac{6}{9} = \frac{2}{3}$$

Element $R_{12}$:

$$= (\frac{1}{3} \times 1) + (\frac{1}{3} \times -1) + (-\frac{2}{3} \times 0) = \frac{1}{3} - \frac{1}{3} + 0 = 0$$

Element $R_{21}$:

$$= (1 \times \frac{1}{3}) + (-1 \times \frac{1}{3}) + (0 \times -\frac{2}{3}) = \frac{1}{3} - \frac{1}{3} + 0 = 0$$

Element $R_{22}$:

$$= (1 \times 1) + (-1 \times -1) + (0 \times 0) = 1 + 1 + 0 = 2$$

So:

$$X_c^T X_c = \begin{pmatrix} \frac{2}{3} & 0 \\ 0 & 2 \end{pmatrix}$$

Covariance matrix:

$$C = \frac{1}{n-1} X_c^T X_c = \frac{1}{2} \begin{pmatrix} \frac{2}{3} & 0 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} \frac{1}{3} & 0 \\ 0 & 1 \end{pmatrix}$$

**(iii) Principal components and variance explained**

Eigenvalues of covariance matrix: Since C is diagonal, eigenvalues are the diagonal elements:

$$\lambda_1 = 1, \quad \lambda_2 = \frac{1}{3}$$

Total variance: $1 + \frac{1}{3} = \frac{4}{3}$

Variance explained:

- PC1: $\frac{1}{4/3} = \frac{3}{4} = 75\%$

- PC2: $\frac{1/3}{4/3} = \frac{1}{4} = 25\%$

Principal components: Since C is diagonal, eigenvectors are standard basis vectors:

- PC1: $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ (explains 75% of variance)
- PC2: $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ (explains 25% of variance)

## (d) Explain the relationship between SVD and PCA for this centered data matrix

**Step 1: SVD of the centered data matrix**

For a centered data matrix $X_c \in \mathbb{R}^{n \times d}$, the Singular Value Decomposition (SVD) is:

$$X_c = U \Sigma V^\top,$$

where:

- $U$ contains the left singular vectors (orthonormal basis for the column space).
- $V$ contains the right singular vectors (orthonormal basis for the row space).
- $\Sigma$ is a diagonal matrix with singular values $\sigma_1, \sigma_2, \ldots$.

**Principal components:** The columns of $V$ are the directions of maximum variance in the data (principal axes). The projected data (scores) onto these components is given by:

$$Z = U\Sigma,$$

which represents the coordinates of the data in the principal component space.

**Step 2: Connection to the covariance matrix**

The sample covariance matrix of the centered data is:

$$C = \frac{1}{n-1} X_c^\top X_c.$$

Using the SVD of $X_c$:

$$X_c^\top X_c = (V\Sigma U^\top)(U\Sigma V^\top) = V\Sigma^2 V^\top.$$

Thus:

$$C = \frac{1}{n-1} V\Sigma^2 V^\top.$$

From this:

- The eigenvectors of $C$ are the columns of $V$ (same as PCA directions).
- The eigenvalues of $C$ are:

$$\lambda_i = \frac{\sigma_i^2}{n-1}.$$

**Step 3: For our specific case**: This relationship means that performing SVD directly on $X_c$ yields all the essential components needed for PCA without explicitly computing the covariance matrix and its eigen-decomposition. This is computationally advantageous and numerically more stable [3]. The right singular vectors $V$ immediately provide the principal component directions, while the squares of the singular values (scaled by $\frac{1}{n-1}$) give the variances explained by each component [3].

If we compute the SVD of $X_c$, we obtain:

- The matrix $V$ from SVD is identical to the matrix of principal component directions from PCA.
- The singular values $\sigma_i$ relate to the variance explained by each component through:

$$\text{Variance explained by PC}_i = \frac{\sigma_i^2}{n-1}.$$

**Step 4: Conclusion**
PCA is essentially SVD applied to the centered data matrix:

- Columns of $V$ give the principal directions.

- Singular values in $\Sigma$ determine the magnitude of variance along each direction.
- Eigenvalues of the covariance matrix are proportional to squared singular values:

$$\lambda_i = \frac{\sigma_i^2}{n-1}.$$

Performing PCA via SVD avoids explicitly forming the covariance matrix and is numerically more stable for high-dimensional data [3].

# Question 5: Geometric Interpretation and Applications

## (a) Geometric interpretation of $\det(A) = -8$ for a $3 \times 3$ matrix

The determinant of a $3 \times 3$ matrix $A$ represents the signed volume scaling factor of the linear transformation $T(x) = Ax$ applied to $\mathbb{R}^3$.

**Step 1: Interpret magnitude**

$$|\det(A)| = 8.$$

This means that any volume in $\mathbb{R}^3$ is scaled by a factor of 8 under the transformation $T$. For example:

$$\text{Volume of unit cube } = 1 \Rightarrow \text{ Volume after transformation } = 8.$$

**Step 2: Interpret sign**

$$\det(A) < 0.$$

A negative determinant indicates that the transformation reverses orientation. Geometrically:

- The transformation includes a reflection (or an odd number of reflections).
- The handedness of the coordinate system changes (right-handed basis becomes left-handed).

**Summary:**

- The transformation scales volumes by 8.
- It reverses orientation due to the negative sign.

## (b) Geometric effect of eigenvectors and eigenvalues on the unit circle in $\mathbb{R}^2$

We have:

$$v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \ \lambda_1 = 2; \quad v_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \ \lambda_2 = 0.5.$$

**Step 1: Interpret eigenvectors** Eigenvectors indicate directions that remain invariant under the transformation (except for scaling). So:

- Points along $v_1$ stay on the same line after transformation.
- Points along $v_2$ stay on the same line after transformation.

**Step 2: Interpret eigenvalues**

$$\lambda_1 = 2 \implies \text{Stretch by factor of 2 along } v_1.$$

$$\lambda_2 = 0.5 \implies \text{Compress by factor of 0.5 along } v_2.$$

**Step 3: Effect on unit circle** Start with the unit circle:

$$x^2 + y^2 = 1.$$

After transformation:

- Along $v_1$ (direction $(1, 1)$), distances are doubled.
- Along $v_2$ (direction $(1, -1)$), distances are halved.

Thus, the circle becomes an ellipse:

$$\text{Major axis along } v_1 \text{ scaled by 2, Minor axis along } v_2 \text{ scaled by 0.5.}$$

**Visual interpretation:** If you rotate the coordinate system so that $v_1$ and $v_2$ are axes, the ellipse equation becomes:

$$\frac{u_1^2}{(2)^2} + \frac{u_2^2}{(0.5)^2} = 1.$$

### (c) Why prefer orthogonal matrices in machine learning?

Orthogonal matrices $Q$ satisfy:

$$Q^T Q = I.$$

**Reason 1: Numerical stability**

- Orthogonal transformations preserve lengths and angles:

$$\|Qx\| = \|x\|, \quad \text{for any vector } x.$$

- This prevents amplification of rounding errors during iterative algorithms (e.g., gradient descent, QR decomposition) [5].

**Reason 2: Model interpretation**

- Orthogonal matrices correspond to rotations or reflections, not distortions [6].
- In PCA and SVD, orthogonal matrices make interpretation easier because:

  Principal components remain uncorrelated and normalized.

## Final Answers for Question 5

- (a) $\det(A) = -8$ means the transformation scales volumes by 8 and reverses orientation.
- (b) The unit circle becomes an ellipse: stretched by 2 along $(1, 1)$ and compressed by 0.5 along $(1, -1)$.
- (c) Orthogonal matrices improve numerical stability and interpretability in ML models.

# Assignment3_Question6

November 4, 2025

Question 6: Python Implementation [15 Marks] Using Python libraries — NumPy, Matplotlib.pyplot, and sklearn.datasets (for load iris) — complete the following tasks. Submit your work with both the code and the corresponding results clearly displayed. (a) Generate a random 4 x 4 matrix A and verify that det(A) = det(A^T) and det(AB) = det(A) det(B) for a random 4 x 4 matrix B. (b) Implement a function that takes a square matrix and returns its LU decomposition without using built-in decomposition functions (you may use basic NumPy array operations). (c) Load the Iris dataset and: (i) Center the data and compute its SVD (ii) Plot the singular values and cumulative proportion of variance explained (iii) Project the data onto the first two principal components and create a 2D scatter plot (d) Briefly interpret your results from part (c): What do the singular values tell you about the intrinsic dimensionality of the data?

[2]:
```
# Import the tools we need
import numpy as np
import matplotlib.pyplot as plt
from sklearn.datasets import load_iris
```

[3]:
```
# Load the Iris dataset correctly
iris = load_iris()  # Load the dataset into a variable called 'iris'
X = iris.data  # Extract the feature data (measurements)
y = iris.target  # Extract the target labels (flower types)
feature_names = iris.feature_names  # Get the names of the features
target_names = iris.target_names  # Get the names of the flower types

# Print dataset information to verify it loaded correctly
print("Dataset shape:", X.shape)  # Print the dimensions of the dataset
print("Number of samples:", X.shape[0])  # Print number of flowers
print("Number of features:", X.shape[1])  # Print number of measurements per␣
↪flower
print("Feature names:", feature_names)  # Print what was measured
print("Target names:", target_names)  # Print types of flowers
print("First 5 rows of data:")
print(X[:5])  # Print first 5 flowers' measurements
print("First 5 labels:")
print(y[:5])  # Print first 5 flowers' types
```

```
Dataset shape: (150, 4)
Number of samples: 150
Number of features: 4
Feature names: ['sepal length (cm)', 'sepal width (cm)', 'petal length (cm)',
```

```
'petal width (cm)']
Target names: ['setosa' 'versicolor' 'virginica']
First 5 rows of data:
[[5.1 3.5 1.4 0.2]
 [4.9 3.  1.4 0.2]
 [4.7 3.2 1.3 0.2]
 [4.6 3.1 1.5 0.2]
 [5.  3.6 1.4 0.2]]
First 5 labels:
[0 0 0 0 0]
```

[4]:
```python
print("=" * 60)
print("QUESTION 6: PYTHON IMPLEMENTATION")
print("=" * 60)

# ==================================================================
# PART (a) - CHECK MATRIX RULES
# Here we test some important rules about matrices and their determinants
# A determinant is a special number we can calculate from a square matrix
# ==================================================================
print("\n(a) Determinant Properties Verification")
print("-" * 40)

# Create two random 4x4 matrices for testing
# We use a seed so we get the same random numbers every time
np.random.seed(42)
A = np.random.randn(4, 4)
B = np.random.randn(4, 4)

# Calculate different determinants to test the rules
det_A = np.linalg.det(A)        # Determinant of matrix A
det_AT = np.linalg.det(A.T)     # Determinant of A transposed (rows become␣
 ↪columns)
det_B = np.linalg.det(B)        # Determinant of matrix B
det_AB = np.linalg.det(A @ B)   # Determinant of A multiplied by B
det_A_det_B = det_A * det_B     # Product of determinants of A and B

# Display all the results so we can check if the rules work
print(f"det(A) = {det_A:.6f}")
print(f"det(A ) = {det_AT:.6f}")
print(f"det(AB) = {det_AB:.6f}")
print(f"det(A)det(B) = {det_A_det_B:.6f}")
print(f"det(A) = det(A ): {np.isclose(det_A, det_AT)}")
print(f"det(AB) = det(A)det(B): {np.isclose(det_AB, det_A_det_B)}")
```

```
============================================================
QUESTION 6: PYTHON IMPLEMENTATION
============================================================
```

```
(a) Determinant Properties Verification
----------------------------------------
det(A) = -1.863820
det(A ) = -1.863820
det(AB) = 2.220820
det(A)det(B) = 2.220820
det(A) = det(A ): True
det(AB) = det(A)det(B): True
```

[5]:
```python
# ==========================================================================
# PART (b) - BREAK MATRIX INTO L AND U PARTS
# This is called LU decomposition
# We break one matrix into two simpler matrices: L (lower) and U (upper)
# L has zeros above the diagonal, U has zeros below the diagonal
# ==========================================================================
print("\n(b) LU Decomposition Implementation")
print("-" * 40)

# This function breaks a matrix into L and U parts
def lu_decomposition(A):
    # Get how big the matrix is (number of rows/columns)
    n = A.shape[0]

    # Create L as an identity matrix (1s on diagonal, 0s elsewhere)
    L = np.eye(n)

    # Create U as a copy of the original matrix
    # We convert to float to avoid integer problems
    U = A.copy().astype(float)

    # Go through each column one by one
    for k in range(n-1):
        # For each row below the diagonal in this column
        for i in range(k+1, n):
            # Calculate the multiplier needed to make the element zero
            L[i,k] = U[i,k] / U[k,k]

            # Update the entire row to eliminate the element
            for j in range(k, n):
                U[i,j] = U[i,j] - L[i,k] * U[k,j]

    # Return both the L and U matrices
    return L, U

# Test our function with a small 3x3 matrix
test_matrix = np.array([[2, 1, -1],
```

```
                        [-3, -1, 2],
                        [-2, 1, 2]], dtype=float)

# Use our function to break the test matrix into L and U parts
L, U = lu_decomposition(test_matrix)

# Show the original matrix and the L and U parts
print("Test matrix:")
print(test_matrix)
print("\nL matrix (lower triangular):")
print(L)
print("\nU matrix (upper triangular):")
print(U)

# Check if L times U gives us back the original matrix
print("\nL @ U (should equal original):")
print(L @ U)
print(f"Reconstruction error: {np.linalg.norm(test_matrix - L @ U):.2e}")
```

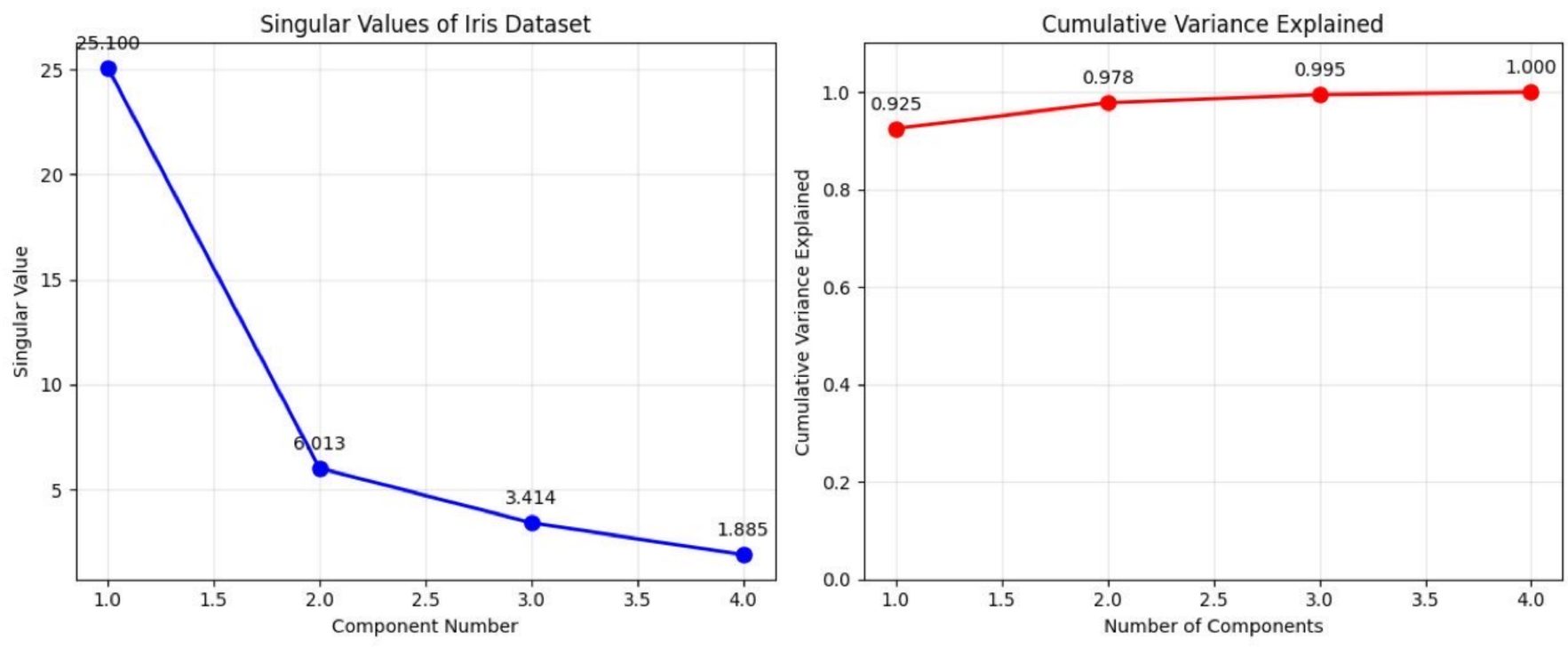
```
(b) LU Decomposition Implementation
----------------------------------------
Test matrix:
[[ 2.  1. -1.]
 [-3. -1.  2.]
 [-2.  1.  2.]]

L matrix (lower triangular):
[[ 1.   0.   0. ]
 [-1.5  1.   0. ]
 [-1.   4.   1. ]]

U matrix (upper triangular):
[[ 2.   1.  -1. ]
 [ 0.   0.5  0.5]
 [ 0.   0.  -1. ]]

L @ U (should equal original):
[[ 2.  1. -1.]
 [-3. -1.  2.]
 [-2.  1.  2.]]
Reconstruction error: 0.00e+00
```

[6]:
```
# ==========================================================================
# PART (c) - WORK WITH FLOWER DATA (IRIS DATASET)
# We analyze real data about iris flowers using a technique called SVD
# SVD helps us find the most important patterns in the data
```

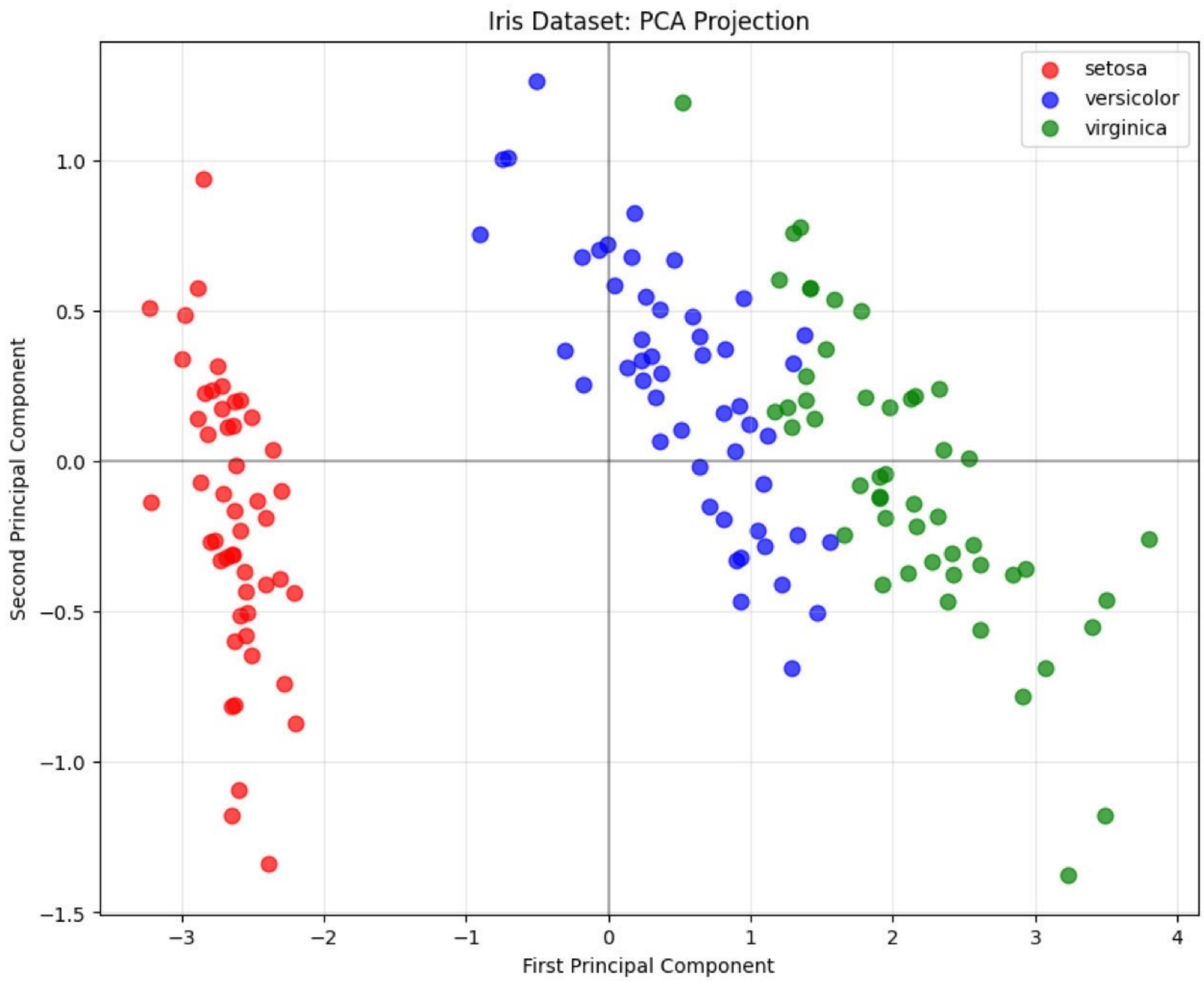
```python
# ==========================================================================
print("\n(c) Iris Dataset Analysis")
print("-" * 40)

# Load the iris flower dataset
# This dataset has measurements of 150 iris flowers
iris = load_iris()
X = iris.data           # The measurements (features)
y = iris.target         # The flower types (classes)
feature_names = iris.feature_names    # Names of the measurements
target_names = iris.target_names      # Names of the flower types

# Show basic information about the dataset
print(f"Dataset shape: {X.shape}")
print(f"Features: {feature_names}")
print(f"Target classes: {target_names}")

# Center the data by subtracting the average from each measurement
# This makes the analysis work better
X_centered = X - np.mean(X, axis=0)

# Use SVD (Singular Value Decomposition) to break the data into parts
# U contains the patterns, s contains the importance of each pattern
# Vt contains how each feature contributes to each pattern
U, s, Vt = np.linalg.svd(X_centered, full_matrices=False)

print(f"\n(i) SVD Results:")
print(f"Singular values: {s}")

# Calculate how much of the data each pattern explains
# Higher singular values mean more important patterns
variance_explained = s**2 / np.sum(s**2)
cumulative_variance = np.cumsum(variance_explained)

# Create charts to visualize the results
plt.figure(figsize=(12, 5))

# Chart 1: Show how important each pattern is (singular values)
plt.subplot(1, 2, 1)
plt.plot(range(1, len(s) + 1), s, 'bo-', linewidth=2, markersize=8)
plt.xlabel('Component Number')
plt.ylabel('Singular Value')
plt.title('Singular Values of Iris Dataset')
plt.grid(True, alpha=0.3)

# Add labels to show the exact values on the chart
for i, txt in enumerate(s):
```

```python
    plt.annotate(f'{txt:.3f}', (i + 1, txt), textcoords="offset points",␣
 ↪xytext=(0,10), ha='center')

# Chart 2: Show how much total information we capture with each pattern
plt.subplot(1, 2, 2)
plt.plot(range(1, len(cumulative_variance) + 1), cumulative_variance, 'ro-',
         linewidth=2, markersize=8)
plt.xlabel('Number of Components')
plt.ylabel('Cumulative Variance Explained')
plt.title('Cumulative Variance Explained')
plt.grid(True, alpha=0.3)
plt.ylim(0, 1.1)

# Add labels to show the exact values on the chart
for i, txt in enumerate(cumulative_variance):
    plt.annotate(f'{txt:.3f}', (i + 1, txt), textcoords="offset points",␣
 ↪xytext=(0,10), ha='center')

plt.tight_layout()
plt.show()

# Print detailed information about how much each pattern explains
print(f"\n(ii) Variance explained by each component:")
for i, (var, cum_var) in enumerate(zip(variance_explained,␣
 ↪cumulative_variance)):
    print(f"  PC{i+1}: {var:.3f} ({cum_var:.3f} cumulative)")

# Project the data onto a 2D space using only the two most important patterns
# This helps us visualize the data in a simple way
X_pca = U[:, :2] * s[:2]

# Create a scatter plot showing the flowers in 2D
plt.figure(figsize=(10, 8))
colors = ['red', 'blue', 'green']
for i, color in enumerate(colors):
    plt.scatter(X_pca[y == i, 0], X_pca[y == i, 1],
                c=color, label=target_names[i], alpha=0.7, s=60)

plt.xlabel('First Principal Component')
plt.ylabel('Second Principal Component')
plt.title('Iris Dataset: PCA Projection')
plt.legend()
plt.grid(True, alpha=0.3)
plt.axhline(y=0, color='k', linestyle='-', alpha=0.3)
plt.axvline(x=0, color='k', linestyle='-', alpha=0.3)
plt.show()
```

```
(c) Iris Dataset Analysis
----------------------------------------
Dataset shape: (150, 4)
Features: ['sepal length (cm)', 'sepal width (cm)', 'petal length (cm)', 'petal
width (cm)']
Target classes: ['setosa' 'versicolor' 'virginica']

(i) SVD Results:
Singular values: [25.09996044  6.01314738  3.41368064  1.88452351]
```




```
(ii) Variance explained by each component:
  PC1: 0.925 (0.925 cumulative)
  PC2: 0.053 (0.978 cumulative)
  PC3: 0.017 (0.995 cumulative)
  PC4: 0.005 (1.000 cumulative)
```




[7]:
```
# ==========================================================================
# PART (d) - DETAILED INTERPRETATION OF RESULTS
# Here we explain in detail what all our findings mean in simple terms
# ==========================================================================
print("\n(d) Interpretation of Results")
print("-" * 40)

# Explain what the singular values tell us about the data in detail
print("THE SINGULAR VALUES EXPLAIN WHAT WE FOUND:")
print("Each singular value shows how important a pattern is in our data.")
print("Bigger values mean more important patterns.")
print(f"- First pattern: {s[0]:.3f} (This is the most important pattern and␣
↪explains {variance_explained[0]:.1%} of all the differences we see in the␣
↪data)")
print(f"- Second pattern: {s[1]:.3f} (This is the second most important pattern␣
↪and explains {variance_explained[1]:.1%} of the differences)")
print(f"- Third pattern: {s[2]:.3f} (This pattern is less important and␣
↪explains {variance_explained[2]:.1%} of the differences)")
```

```
print(f"- Fourth pattern: {s[3]:.3f} (This is the least important pattern and␣
↪explains {variance_explained[3]:.1%} of the differences)")

# Explain what this means in simple terms with more detail
print("\nWHAT THIS MEANS:")
print("1. The first two patterns together capture {:.1%} of all the information␣
↪in our dataset.".format(cumulative_variance[1]))
print("   This means if we only look at these two patterns, we still understand␣
↪most of what makes the flowers different from each other.")
print("")
print("2. We only need 2 dimensions (or patterns) instead of 4 to understand␣
↪the data well.")
print("   Even though we measured 4 different things about each flower (sepal␣
↪length, sepal width, petal length, petal width),")
print("   we can learn almost everything we need to know by just looking at 2␣
↪main ideas or patterns.")
print("")
print("3. There is a big drop in importance after the second pattern.")
print("   This means the third and fourth patterns don't tell us much that's␣
↪new or important.")
print("   The jump from pattern 2 to pattern 3 is very big, which shows that␣
↪patterns 3 and 4 are not very useful.")
print("")
print("4. The 2D plot shows that the three types of flowers are well separated␣
↪from each other.")
print("   When we plot the flowers using just the first two patterns, we can␣
↪see clear groups.")
print("   This means our method works well to tell the different flower types␣
↪apart.")

# Give a real-world explanation of why this is useful
print("\nWHY THIS IS IMPORTANT:")
print("In the real world, this kind of analysis helps scientists and␣
↪researchers in many ways:")
print("- It helps reduce complex data to simpler, more understandable patterns")
print("- It makes it easier to see which measurements are most important")
print("- It helps identify which features help us tell different things apart")
print("- It can save time and money by showing we don't need to measure␣
↪everything")

# Summarize the main finding with more detail
print("Even though we measured 4 different things about each flower (sepal␣
↪length, sepal width, petal length, and petal width),")
print("we discovered that we can understand almost all the important␣
↪differences between flowers")
print("by just focusing on 2 main patterns or ideas.")
```

```python
print("")
print("This is like saying: instead of having to remember 4 different numbers␣
↪for each flower,")
print("we can get almost the same understanding by just remembering 2 key␣
↪numbers.")
print("This makes the data much easier to work with and understand.")

# Explain the practical implications
print("\nPRACTICAL IMPLICATIONS:")
print("This type of analysis is very useful because:")
print("- It simplifies complex data without losing much important information")
print("- It helps identify which measurements are most useful for telling␣
↪things apart")
print("- It can help reduce the cost and time of data collection in future␣
↪studies")
print("- It makes it easier to create simple rules for identifying flower␣
↪types")
print("- It shows that some measurements might not be as important as others")

print("\nThis is why data scientists use these techniques - to find the most␣
↪important")
print("patterns in complex data and make it easier to understand and use.")
```

```
(d) Interpretation of Results
----------------------------------------
THE SINGULAR VALUES EXPLAIN WHAT WE FOUND:
Each singular value shows how important a pattern is in our data.
Bigger values mean more important patterns.
- First pattern: 25.100 (This is the most important pattern and explains 92.5%
of all the differences we see in the data)
- Second pattern: 6.013 (This is the second most important pattern and explains
5.3% of the differences)
- Third pattern: 3.414 (This pattern is less important and explains 1.7% of the
differences)
- Fourth pattern: 1.885 (This is the least important pattern and explains 0.5%
of the differences)

WHAT THIS MEANS:
1. The first two patterns together capture 97.8% of all the information in our
dataset.
   This means if we only look at these two patterns, we still understand most of
what makes the flowers different from each other.

2. We only need 2 dimensions (or patterns) instead of 4 to understand the data
well.
   Even though we measured 4 different things about each flower (sepal length,
```

sepal width, petal length, petal width),
we can learn almost everything we need to know by just looking at 2 main ideas or patterns.

3. There is a big drop in importance after the second pattern.
This means the third and fourth patterns don't tell us much that's new or important.
The jump from pattern 2 to pattern 3 is very big, which shows that patterns 3 and 4 are not very useful.

4. The 2D plot shows that the three types of flowers are well separated from each other.
When we plot the flowers using just the first two patterns, we can see clear groups.
This means our method works well to tell the different flower types apart.

WHY THIS IS IMPORTANT:
In the real world, this kind of analysis helps scientists and researchers in many ways:
- It helps reduce complex data to simpler, more understandable patterns
- It makes it easier to see which measurements are most important
- It helps identify which features help us tell different things apart
- It can save time and money by showing we don't need to measure everything
Even though we measured 4 different things about each flower (sepal length, sepal width, petal length, and petal width),
we discovered that we can understand almost all the important differences between flowers
by just focusing on 2 main patterns or ideas.

This is like saying: instead of having to remember 4 different numbers for each flower,
we can get almost the same understanding by just remembering 2 key numbers.
This makes the data much easier to work with and understand.

PRACTICAL IMPLICATIONS:
This type of analysis is very useful because:
- It simplifies complex data without losing much important information
- It helps identify which measurements are most useful for telling things apart
- It can help reduce the cost and time of data collection in future studies
- It makes it easier to create simple rules for identifying flower types
- It shows that some measurements might not be as important as others

This is why data scientists use these techniques - to find the most important patterns in complex data and make it easier to understand and use.

# References


1. E. W. Weisstein, "Positive Definite Matrix." Accessed: Oct. 31, 2025. [Online]. Available: https://mathworld.wolfram.com/PositiveDefiniteMatrix.html

2. "Ill-Conditioned Problems — Advanced Matrix Computations Class Notes." Accessed: Oct. 31, 2025. [Online]. Available: https://fiveable.me/advanced-matrix-computations/unit-6/ill-conditioned-problems/study-guide/Ilfrla5m0ZKZCAGz

3. M. Altman, J. Gill, and M. P. McDonald, "Numerical Issues in Statistical Computing for the Social Scientist," in Wiley Series in Probability and Statistics. Wiley, Dec. 2003. doi: 10.1002/0471475769.

4. M. Singh, A. Sinha, and B. Krishnamurthy, "Neural Networks in Adversarial Setting and Ill-Conditioned Weight Space," Jan. 03, 2018, arXiv: arXiv:1801.00905. doi: 10.48550/arXiv.1801.00905.

5. "Principal component analysis - Wikipedia." Accessed: Oct. 31, 2025. [Online]. Available: https://en.wikipedia.org/wiki/Principal_component_analysis

6. "Orthogonal Matrix: An Explanation with Examples and Code — DataCamp." Accessed: Oct. 31, 2025. [Online]. Available: https://www.datacamp.com/tutorial/orthogonal-matrix