Q-2

(vi) An orthonormal basis in the context of PCA means that the principal components [PCs] are both orthogonal [perpendicular to each other] and normalized [have unit length].

Given a dataset X with n observations & d features, [Assume data is centered], the covariance vector C is:

$$C = \frac{1}{n-1} X^T X$$

Eigen composition of C:

$$C = V \Lambda V^T$$

V = eigenvectors of C, each column of V, v_i is an eigenvector

Λ: diagonal ~~vector~~ matrix whose diagonal elements are eigenvalues λ_i of C.

V^T: transpose of V.

→ Any two eigenvectors are orthogonal.
v_i & v_j of C are orthogonal.

$$v_i^T \cdot v_j = 0 \qquad i \neq j$$

→ They can be normalized to unit length.

$\|v_i\| = 1$, $\|v_i\|$ is norm [euclidean]

→ For v_i & v_j [eigenvectors] with eigenvalues λ_i, λ_j, the covariance mat [C],

$$C v_i = \lambda_i v_i$$
$$C v_j = \lambda_j v_j$$

Dot product with v_j.

$\cancel{C v_j v_i = \lambda}$

$$v_j^T C v_i = \lambda_i v_j^T v_i$$

C is symmetric.

$$\therefore \quad v_j^T C v_i = (C v_j)^T v_i$$
$$= \lambda_j v_j^T v_i$$

$$\therefore \quad \lambda_i v_j^T v_i = \lambda_j v_j^T v_i$$

If $\lambda_1 \neq \lambda_2$, $\therefore \boxed{v_j^T v_i = 0} \rightarrow$ Orthogonality.

→ Normalization.

Each eigenvector can be scaled such that $||v_i|| = 1$.

$v_i' = \frac{v_i}{||v_i||}$, v_i' has unit length.

Thus, PCA naturally forms an orthonormal basis where each principal component is represented by an eigenvector of the covariance matrix. This orthonormal basis optimally reduces the dimensionality of the data by retaining the most variance in fewest principal components.