

## ① Steps involved in PCA :-

### ① Represent data into structure -

- ↳ Represent dataset into a structure.
- ↳ Represent 2D matrix of independent variable  $x$ .
- ↳ Each row = data items & columns = Features.
- ↳ Num. of columns is dimension of dataset.

### ② Standardize the Dataset -

- ↳ In a column, high variance features more important than low variance.
- ↳ If feature importance variance independent, then divide data item in column with std. deviation of column.
- ↳ Name the matrix  $Z$ .

### ③ Calculate Covariance<sup>Matrix</sup> of $Z$ -

- ↳ Covariance measures how much 2 variables  $\Rightarrow$  change in relation to each other.

- ↳ Formula for variance:

$$\text{cov}(x_1, x_2) = \sum_{i=1}^N \frac{(x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2)}{N-1}$$

- ↳ Value can be +ve, -ve OR zeros.

- Positive:  $x_1$  directly proportional to  $x_2$ .
- Negative:  $x_1$  inversely proportional to  $x_2$ .
- Zeros: No direct relation.

#### ④ Compute Eigenvalues & Eigenvectors -

↳ Let  $A$  be square  $n \times n$  matrix &  $x$  be non-zero vectors for which:

$$Ax = \lambda x$$

for some scalar values  $\lambda$ .

$\lambda$  - Eigenvalue of matrix  $A$ .

$x$  - Eigenvector of matrix  $A$ .

↳ Find Eigenvalues -

$$C - \lambda I = 0$$

↳ Returns n num of eigenvalues.

↳ Find Eigenvector - For each eigenvalues.

$$Cz = \lambda z$$

↳ Eigen values

↳ Eigenvectors (To be find).

#### ⑤ Sort the Eigen vectors -

↳ Sort all eigenvalues in decreasing order from largest to smallest.

↳ Simultaneously sort eigenvectors according to matrix of eigenvalues.

↳ Resultant matrix named  $P^*$ .



⑥ Calculate new features / Principal components (PC) -

↳ Multiply  $P^*$  matrix to  $Z$ .

$$Z^* = P^* \times Z$$

↳ In  $Z^*$ , each observation is linear combination of original features.

↳ Each  $Z^*$  column independent of each other.

⑦ Remove less imp / unimportant features -

↳ From new feature set, decide which features to keep & what to remove.

↳ Only keep relevant / important features & remove unimportant features.