PCA Steps

1. Data of 'n' observations
2. No. of features = 'p'
3. Form a matrix of size 'n x p' with deviations from mean for each of the variables
4. Calculate covariance matrix. $(p \times p)$

$$cov(x,y) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{(n-1)}$$

(5) Calculate eigen values & eigenvectors of the covariance matrix

(6) Choose principal components & form a new feature vector.

(7) Derive the new data set

Trace (S) : The fraction of the total variance accounted for by the jth principal component is :

$$\frac{T_j}{Trace(S)}$$

PCA

Where $\text{Trace}(S) = \sum T_j$

$$\text{Trace}(S) = 322.4 + 27.7$$
$$= 350.1$$

Total variance accounted by 1st principal component is:

$$\frac{T_1}{\text{Trace}(S)} = \frac{322.3}{350.1} = 0.92$$

i.e. 92% of total system variance is reflected by 1st PC.

the remaining 8% is represented
by 2nd component.

Hence PC 2 is neglected &
only PC 1 is considered.

From the 2 eigen vectors the
feature vector is selected:

$$\begin{bmatrix} 0.801 \\ 0.599 \end{bmatrix}$$

Last step: multiply the dev.
cols with
the above selected feature
vector: Dev. vec 1 = 15 rows
" 2 = 15 rows.

$\therefore$ dev mahi = 15×2

feature vech = 2×1

$\therefore$ o/p = 15×1

$$\boxed{Z = X \cdot A}$$

Multiply the matrix of deviations from mean with the eigen vectors to form a reduced vector.

This will be the reduced feature set.

In MLR:

→ u have p independent vars.

→ u formulate 'p' principal component

→

$$y = f(x_1, x_2 \ldots, x_p)$$

⇒ we set 'p' principal component

Each 'p' contributes to variance in 'y'.

choose 'q' of 'p' principal components (q<p)

to be included in the Regression eqn.