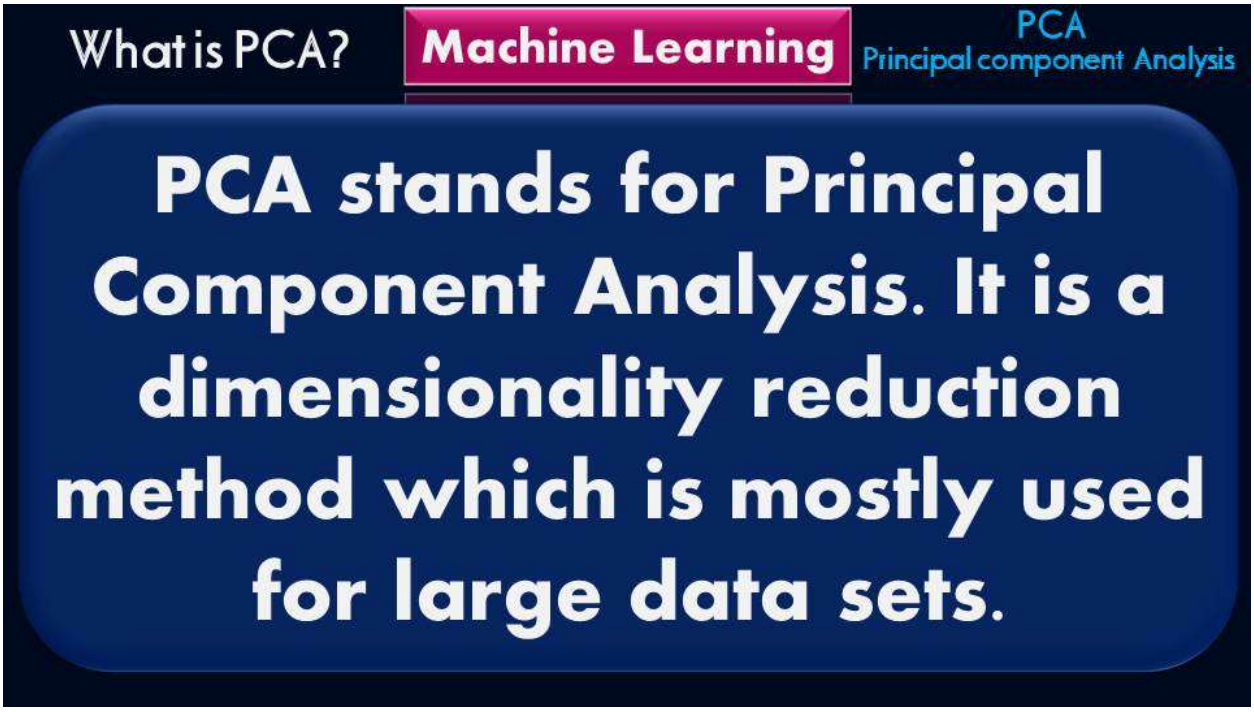
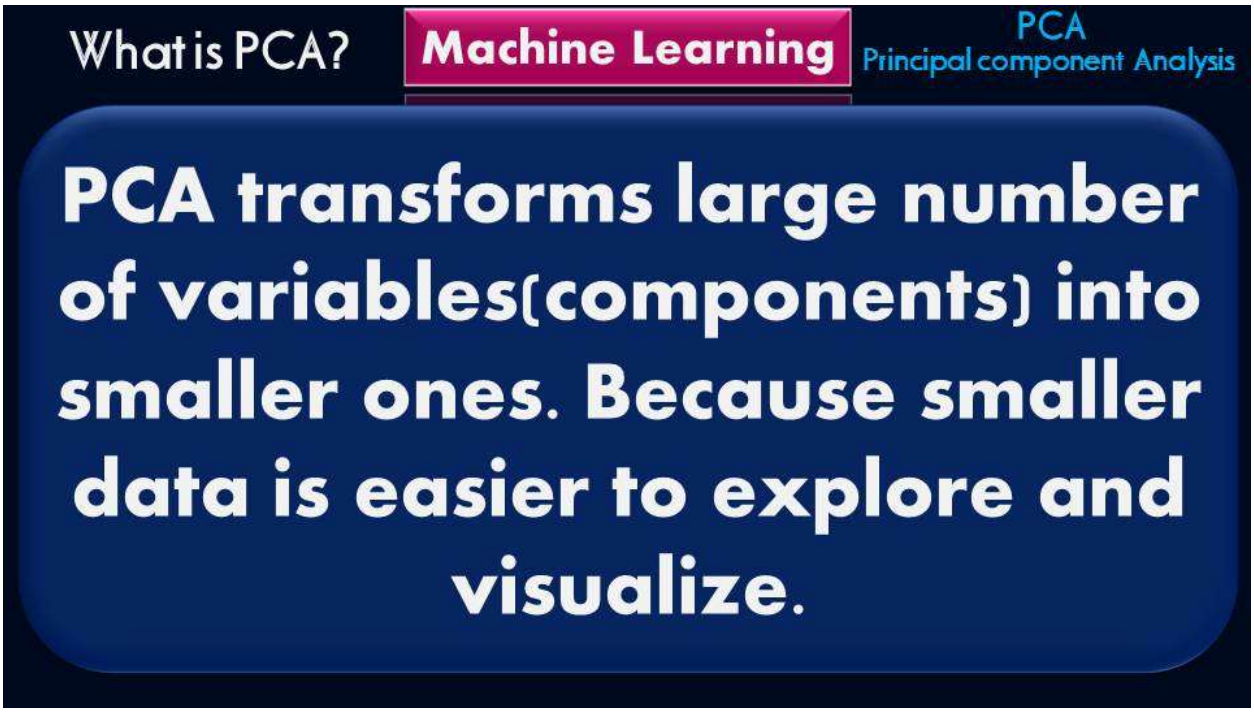
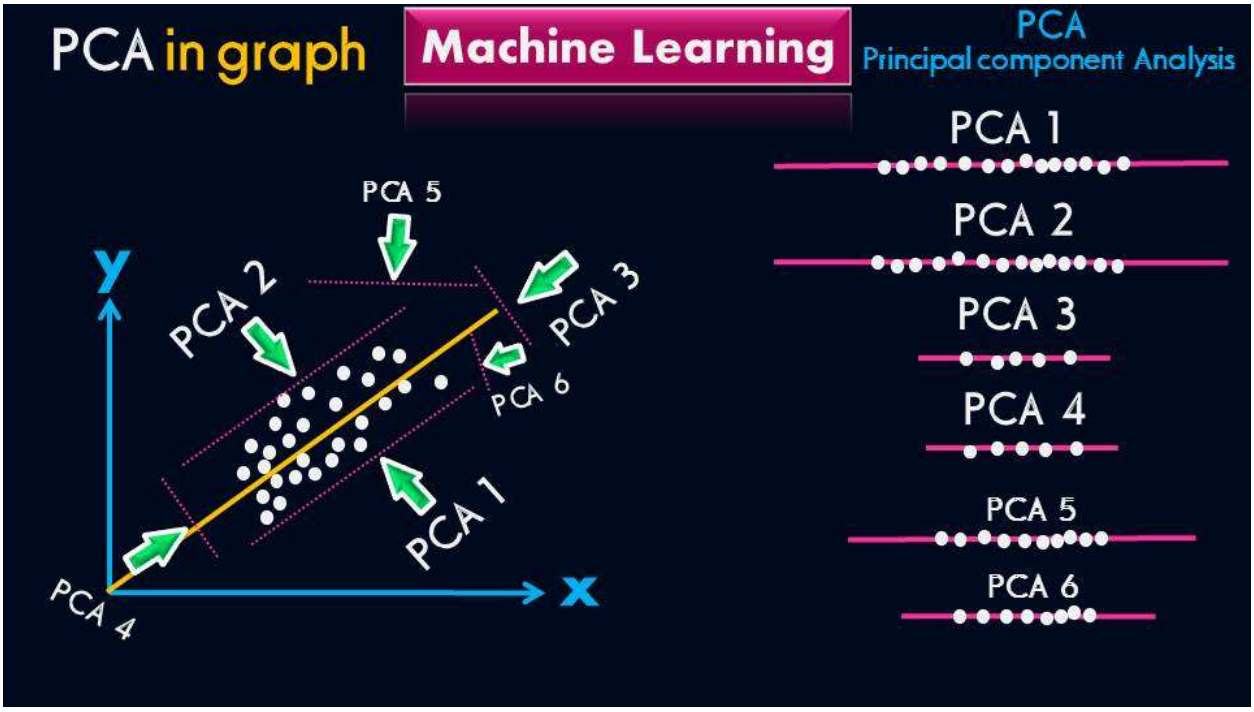
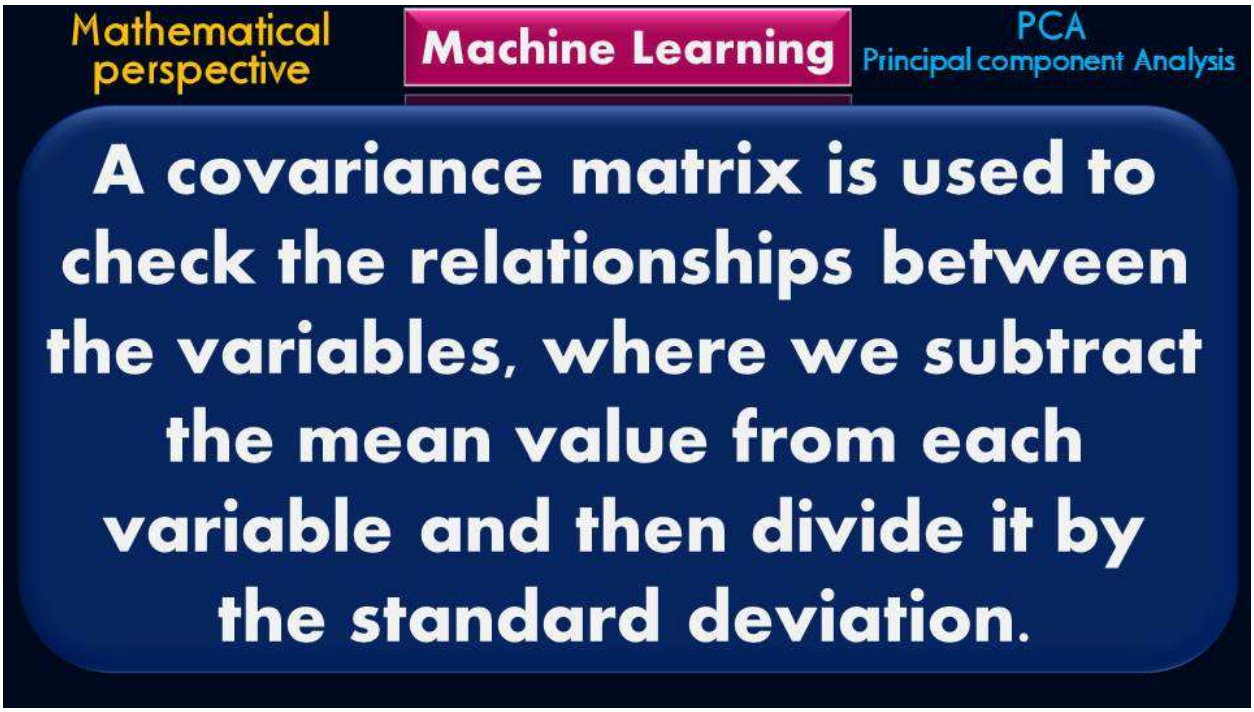
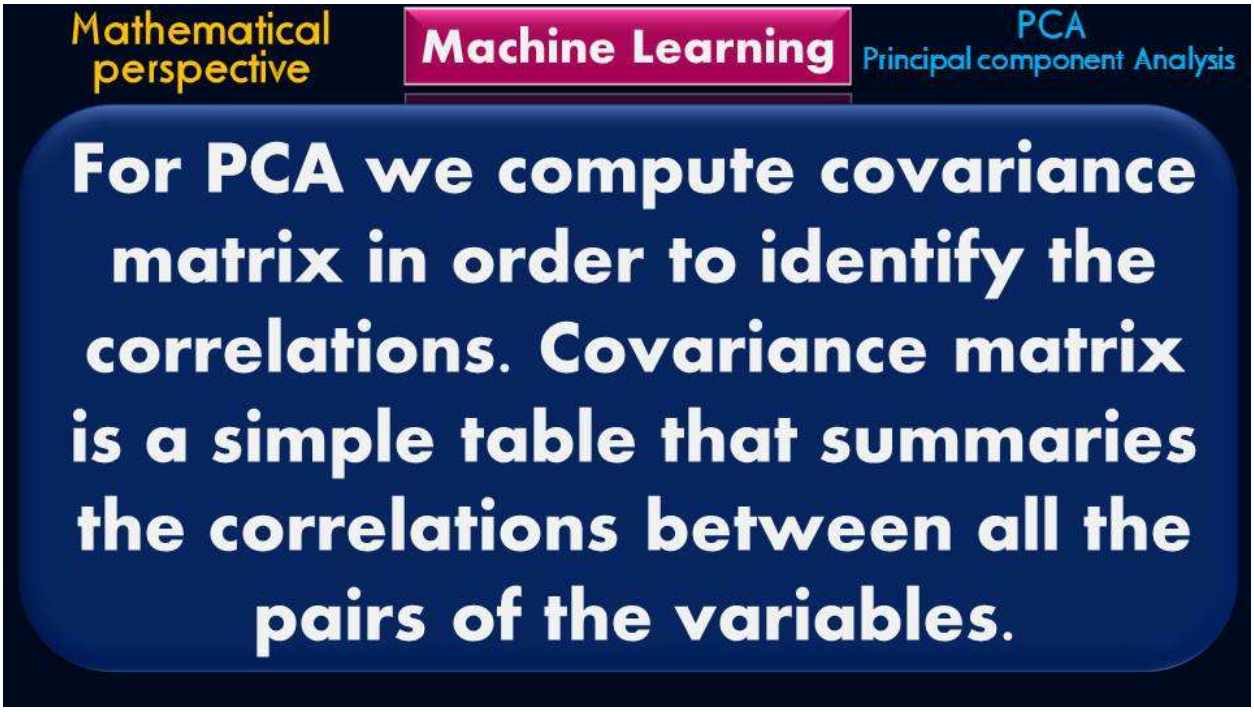
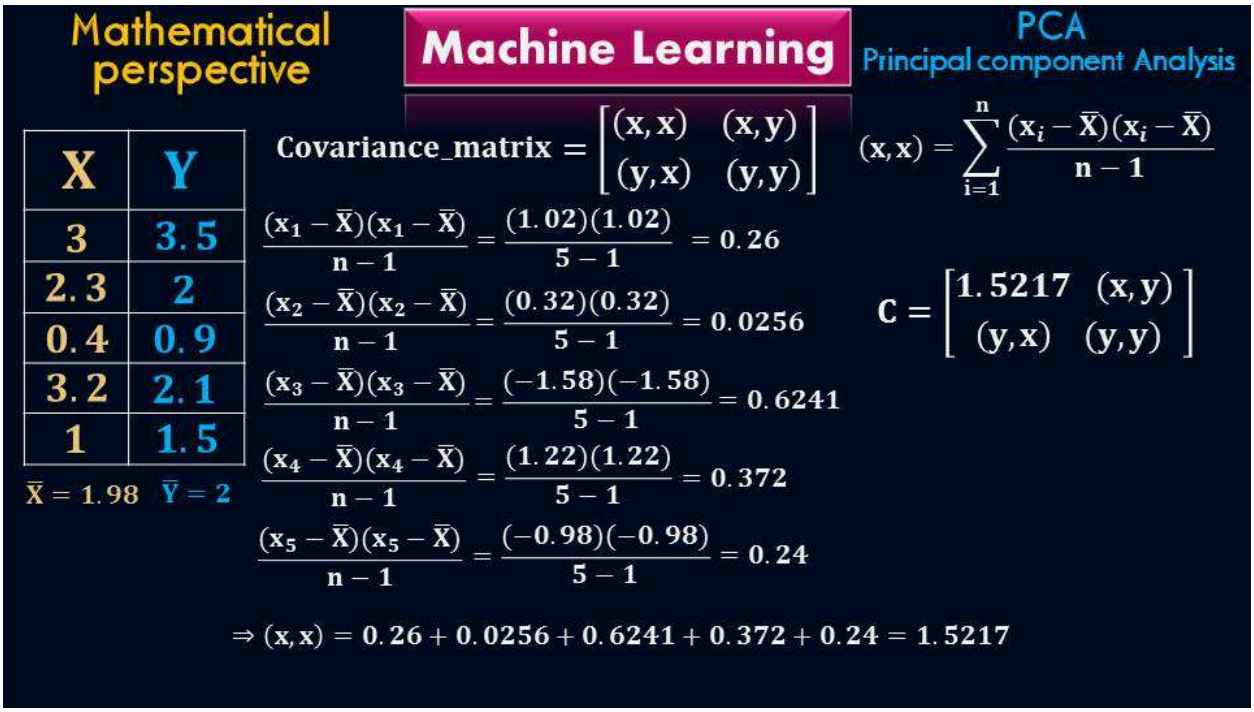

What is PCA?
Machine Learning
PCA
Principal component Analysis
PCA stands for Principal Component Analysis. It is a dimensionality reduction method which is mostly used for large data sets.

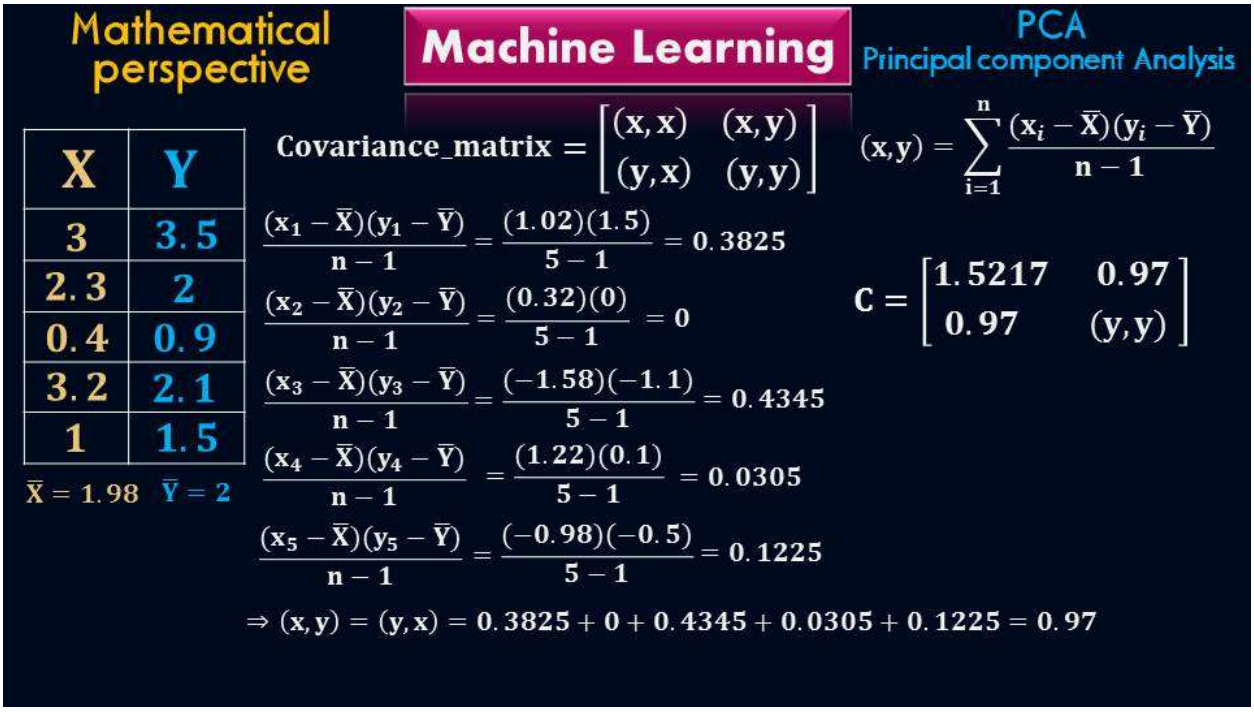

What is PCA?
Machine Learning
PCA
Principal component Analysis
PCA transforms large number of variables(components) into smaller ones. Because smaller data is easier to explore and visualize.

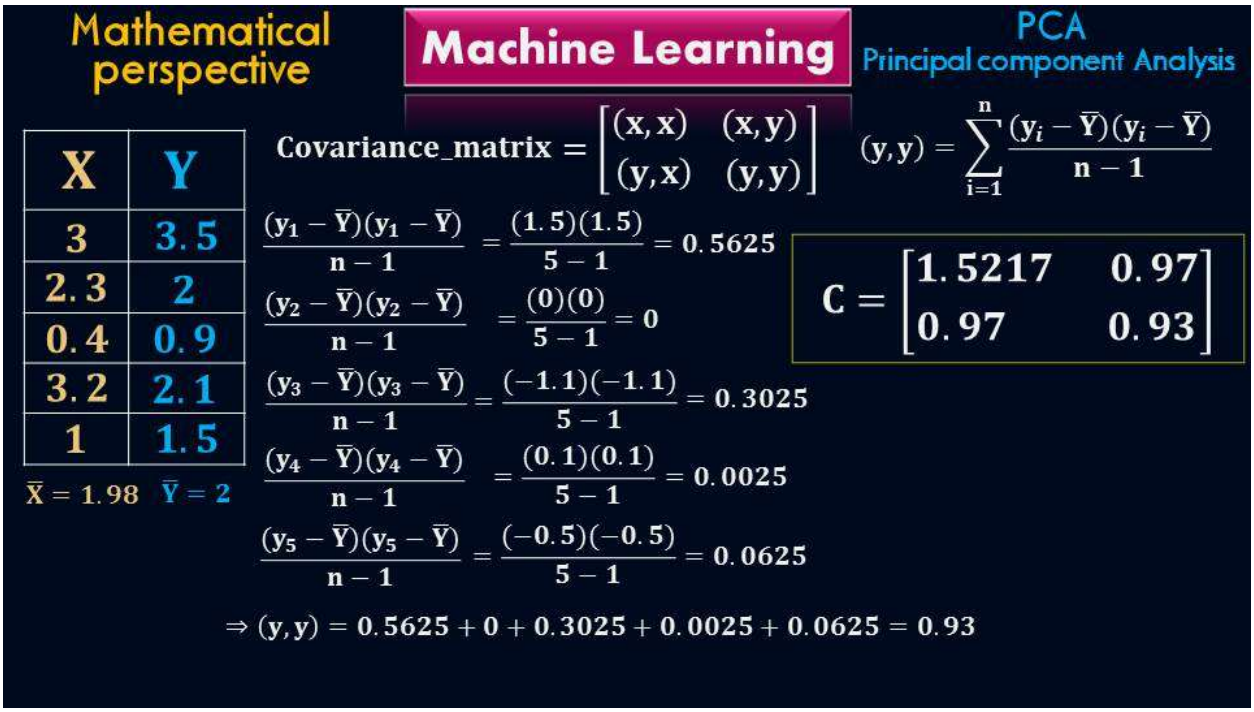

PCA in graph
Machine Learning
PCA
Principal component Analysis
PCA 1
PCA 2
PCA 3
PCA 4
PCA 5
PCA 6
y
x


Mathematical perspective
Machine Learning
PCA
Principal component Analysis
A covariance matrix is used to check the relationships between the variables, where we subtract the mean value from each variable and then divide it by the standard deviation.


Mathematical perspective
Machine Learning
PCA
Principal component Analysis
For PCA we compute covariance matrix in order to identify the correlations. Covariance matrix is a simple table that summaries the correlations between all the pairs of the variables.


Mathematical perspective
Machine Learning
PCA
Principal component Analysis
X Y
3 3.5
2.3 2
0.4 0.9
3.2 2.1
1 1.5
$\bar{X} = 1.98$ $\bar{Y} = 2$
$\text{Covariance\_matrix} = \begin{bmatrix} (x,x) & (x,y) \\ (y,x) & (y,y) \end{bmatrix}$
$(x,x) = \sum_{i=1}^{n} \frac{(x_i - \bar{X})(x_i - \bar{X})}{n-1}$
$\frac{(x_1 - \bar{X})(x_1 - \bar{X})}{n-1} = \frac{(1.02)(1.02)}{5-1} = 0.26$
$\frac{(x_2 - \bar{X})(x_2 - \bar{X})}{n-1} = \frac{(0.32)(0.32)}{5-1} = 0.0256$
$\frac{(x_3 - \bar{X})(x_3 - \bar{X})}{n-1} = \frac{(-1.58)(-1.58)}{5-1} = 0.6241$
$\frac{(x_4 - \bar{X})(x_4 - \bar{X})}{n-1} = \frac{(1.22)(1.22)}{5-1} = 0.372$
$\frac{(x_5 - \bar{X})(x_5 - \bar{X})}{n-1} = \frac{(-0.98)(-0.98)}{5-1} = 0.24$
$C = \begin{bmatrix} 1.5217 & (x,y) \\ (y,x) & (y,y) \end{bmatrix}$
$\Rightarrow (x,x) = 0.26 + 0.0256 + 0.6241 + 0.372 + 0.24 = 1.5217$


Mathematical perspective
Machine Learning
PCA
Principal component Analysis
X Y
3 3.5
2.3 2
0.4 0.9
3.2 2.1
1 1.5
$\bar{X} = 1.98$ $\bar{Y} = 2$
$\text{Covariance\_matrix} = \begin{bmatrix} (x,x) & (x,y) \\ (y,x) & (y,y) \end{bmatrix}$
$(x,y) = \sum_{i=1}^{n} \frac{(x_i - \bar{X})(y_i - \bar{Y})}{n-1}$
$\frac{(x_1 - \bar{X})(y_1 - \bar{Y})}{n-1} = \frac{(1.02)(1.5)}{5-1} = 0.3825$
$\frac{(x_2 - \bar{X})(y_2 - \bar{Y})}{n-1} = \frac{(0.32)(0)}{5-1} = 0$
$\frac{(x_3 - \bar{X})(y_3 - \bar{Y})}{n-1} = \frac{(-1.58)(-1.1)}{5-1} = 0.4345$
$\frac{(x_4 - \bar{X})(y_4 - \bar{Y})}{n-1} = \frac{(1.22)(0.1)}{5-1} = 0.0305$
$\frac{(x_5 - \bar{X})(y_5 - \bar{Y})}{n-1} = \frac{(-0.98)(-0.5)}{5-1} = 0.1225$
$C = \begin{bmatrix} 1.5217 & 0.97 \\ 0.97 & (y,y) \end{bmatrix}$
$\Rightarrow (x,y) = (y,x) = 0.3825 + 0 + 0.4345 + 0.0305 + 0.1225 = 0.97$


Mathematical perspective
Machine Learning
PCA
Principal component Analysis
X Y
3 3.5
2.3 2
0.4 0.9
3.2 2.1
1 1.5
$\bar{X} = 1.98$ $\bar{Y} = 2$
$\text{Covariance\_matrix} = \begin{bmatrix} (x,x) & (x,y) \\ (y,x) & (y,y) \end{bmatrix}$
$(y,y) = \sum_{i=1}^{n} \frac{(y_i - \bar{Y})(y_i - \bar{Y})}{n-1}$
$\frac{(y_1 - \bar{Y})(y_1 - \bar{Y})}{n-1} = \frac{(1.5)(1.5)}{5-1} = 0.5625$
$\frac{(y_2 - \bar{Y})(y_2 - \bar{Y})}{n-1} = \frac{(0)(0)}{5-1} = 0$
$\frac{(y_3 - \bar{Y})(y_3 - \bar{Y})}{n-1} = \frac{(-1.1)(-1.1)}{5-1} = 0.3025$
$\frac{(y_4 - \bar{Y})(y_4 - \bar{Y})}{n-1} = \frac{(0.1)(0.1)}{5-1} = 0.0025$
$\frac{(y_5 - \bar{Y})(y_5 - \bar{Y})}{n-1} = \frac{(-0.5)(-0.5)}{5-1} = 0.0625$
$C = \begin{bmatrix} 1.5217 & 0.97 \\ 0.97 & 0.93 \end{bmatrix}$
$\Rightarrow (y,y) = 0.5625 + 0 + 0.3025 + 0.0025 + 0.0625 = 0.93$