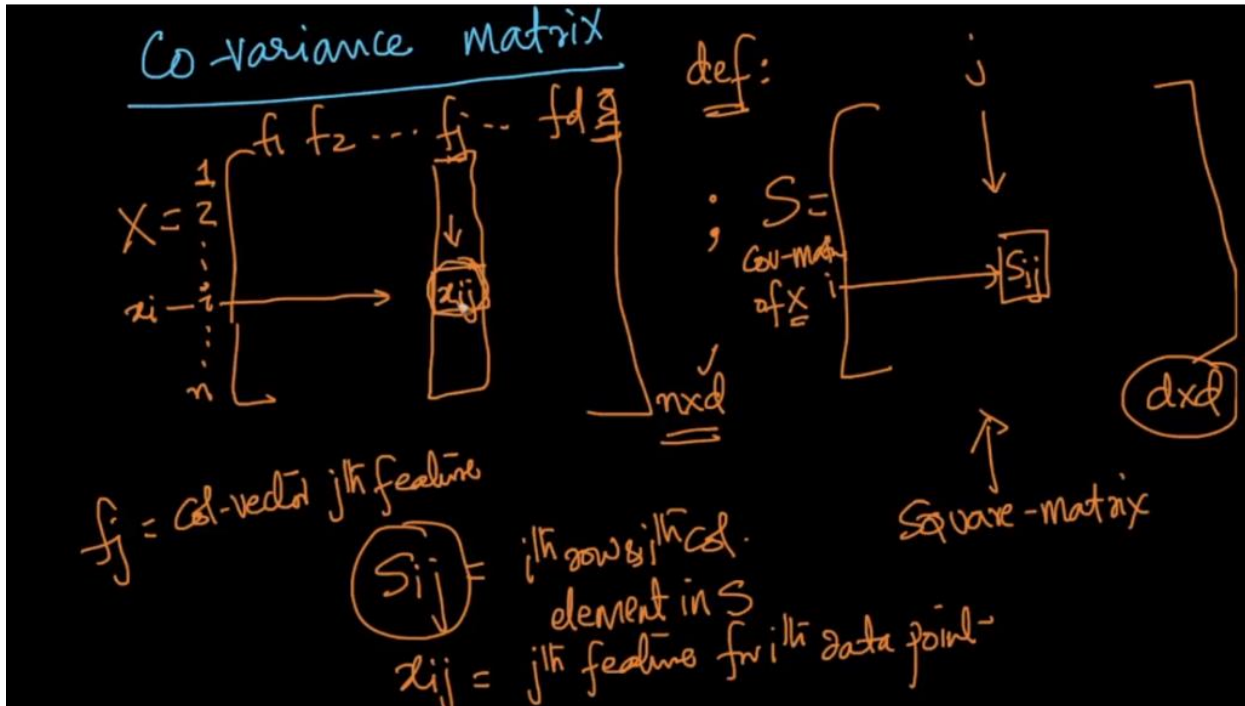


This is one of the interesting data preprocessing in which we find the covariance of each feature with another feature.

Therefore for a data matrix of dimensions $n \times d$, the **co-variance** matrix will be of dimension $d \times d$, because we plot each feature against each feature and therefore co-variance matrix will be a square matrix.



Below figure shows two properties of covariance:

1. If we are finding covariance of same RV X , then it will be equal to variance.
Or we can say if we find the co-variance of one feature with itself then it will be equal to variance of that feature.
2. Covariance (SL, PL) = Covariance (PL, SL)

$$S_{ij} = \text{cov}(f_i, f_j)$$

$i: 1 \rightarrow d$
 $j: 1 \rightarrow d$

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

$$\text{cov}(f_i, f_i) = \text{Var}(f_i)$$

$$\checkmark \text{cov}(x, x) = \text{Var}(x)$$

$$\text{cov}(f_i, f_j) = \text{cov}(f_j, f_i)$$

Below image shows that in co-variance matrix the diagonal will be the co-variance of a feature with itself, hence diagonal represent the variance of the features.

Co-variance matrix is also called symmetric matrix because in symmetric matrix $A_{21} = A_{12}$.

Variances of features

Sq. Symm matrix

$S =$

S_{11} S_{22} S_{33} S_{44} \dots S_{dd}

S_{ij}

S_{ji}

Symm - matrix

Sq. matrix $\begin{cases} d: \text{rows} \\ d: \text{col} \end{cases}$

$d \times d$

Symmetric matrix

$A =$

$A_{32} = S = A_{23}$

$A_{21} = A_{12}$

$A_{ij} = A_{ji} \forall i, j$

3×3

For Finding Co-variance of any data matrix first we'll perform **column standardization**, to bring the mean of all feature to 0.

Since after standardization mean becomes 0, so we'll remove mean from co-variance formula,

Hence now for co-variance of two features we simply multiply their corresponding points and sum up and then we divide by n.

And therefore we can also say that covariance of two features f_1 and f_2 will be dot product as:

$$(f_1 \cdot f_2) / n$$

The image shows handwritten notes on a blackboard. At the top left, a matrix X is defined as $X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1d} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nd} \end{bmatrix}$ with dimensions $n \times d$. Arrows point from the columns to labels f_1, f_2, \dots, f_d . To the right, the matrix is shown as $X = \begin{bmatrix} f_1 & f_2 & \dots & f_d \end{bmatrix}$. Below this, it says "Let (X) Col. Standardized \Rightarrow mean $\{f_i\} = 0$ and std-dev $\{f_i\} = 1$ ". There are handwritten notes "avg" and "pL" under "Col. Standardized". Below this, the covariance formula is written as
$$\text{Cov}(f_1, f_2) = \frac{1}{n} \sum_{i=1}^n (x_{i1} - \mu_1)(x_{i2} - \mu_2)$$
 with arrows pointing from μ_1 to "mean(f_1)" and from μ_2 to "mean(f_2)". There are also handwritten notes "pL" and "pW" near the summation.

$$\text{Cov}(f_1, f_2) = \frac{1}{n} \sum_{i=1}^n x_{i1} * x_{i2}$$

Diagram illustrating the data matrix X with columns f_1 and f_2 . The rows are indexed 1 to n . The element x_{i1} is circled in the f_1 column, and x_{i2} is circled in the f_2 column. A bracket indicates the product $x_{i1} * x_{i2}$ for each row i .

$$\text{Cov}(f_1, f_2) = \frac{1}{n} (f_1^T f_2)$$

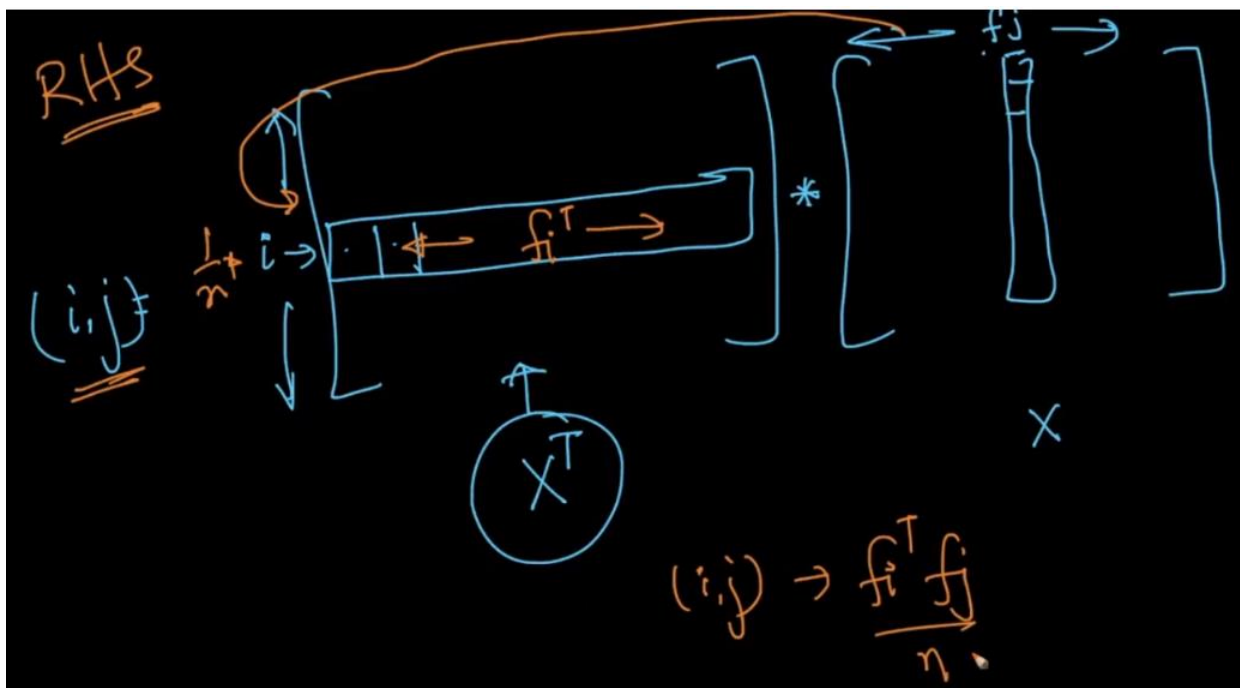
Since after standardization mean becomes 0, and we are making co-variance matrix just by using value of data-matrix, therefore we can also say the co-variance matrix can be $(X^T * X)/n$

$$S_{d \times d} = \frac{1}{n} (X^T)_{d \times n} (X)_{n \times d} = (d \times d) \checkmark$$

data-matrix

(*) assuming X has been col. std

(LHS) $S_{ij} = \text{Cov}(f_i, f_j) = \frac{f_i^T f_j}{n}$



$$S_{d \times d} = X_{d \times n}^T X_{n \times d}$$

if X has been col. std

Notes:

- In co-variance, it's impacted by scale or units of measurement, but in calculating co-variance matrix we are standardizing before finding co-variance, that means we are removing the impact of scaling so therefore in such case co-variance is similar to co-relation.

Covariance is nothing but a measure of correlation. On the contrary, correlation refers to the scaled form of covariance.

The value of correlation takes place between -1 and +1. Conversely, the value of covariance lies between -? and +?. So, if the random variables are standardized before calculating the covariance then covariance is equal to the correlation and has a value between -1 and +1.

