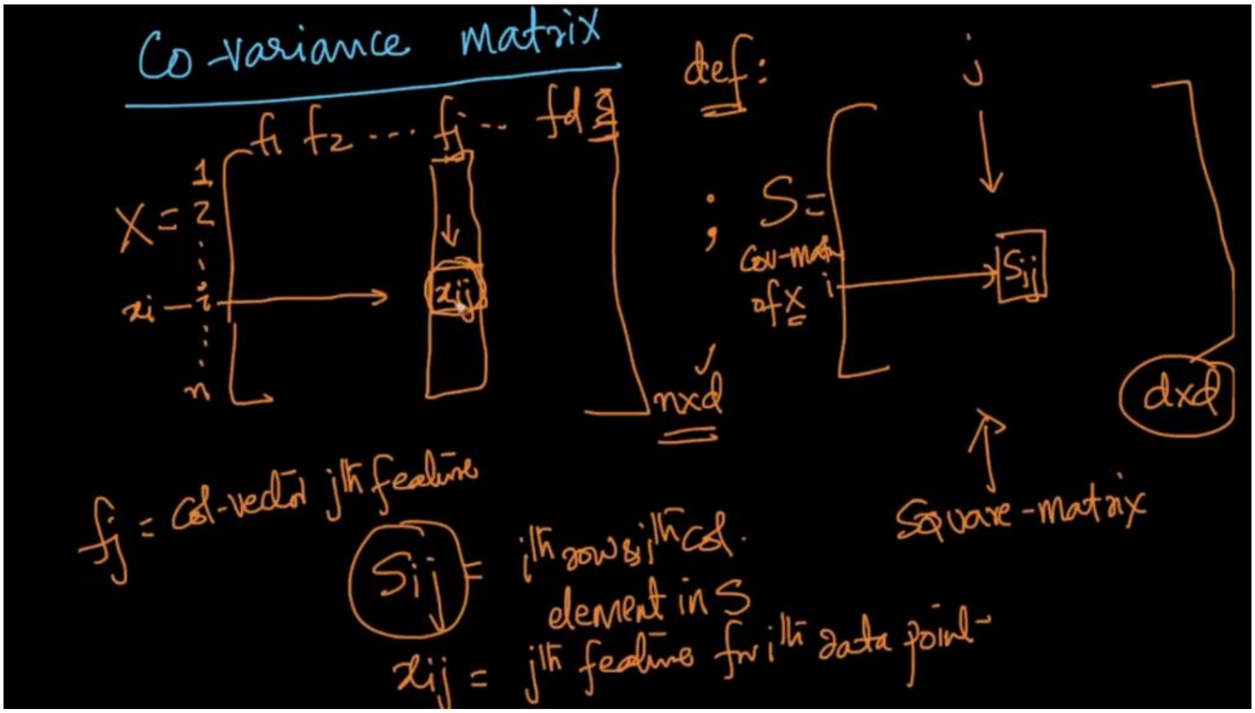
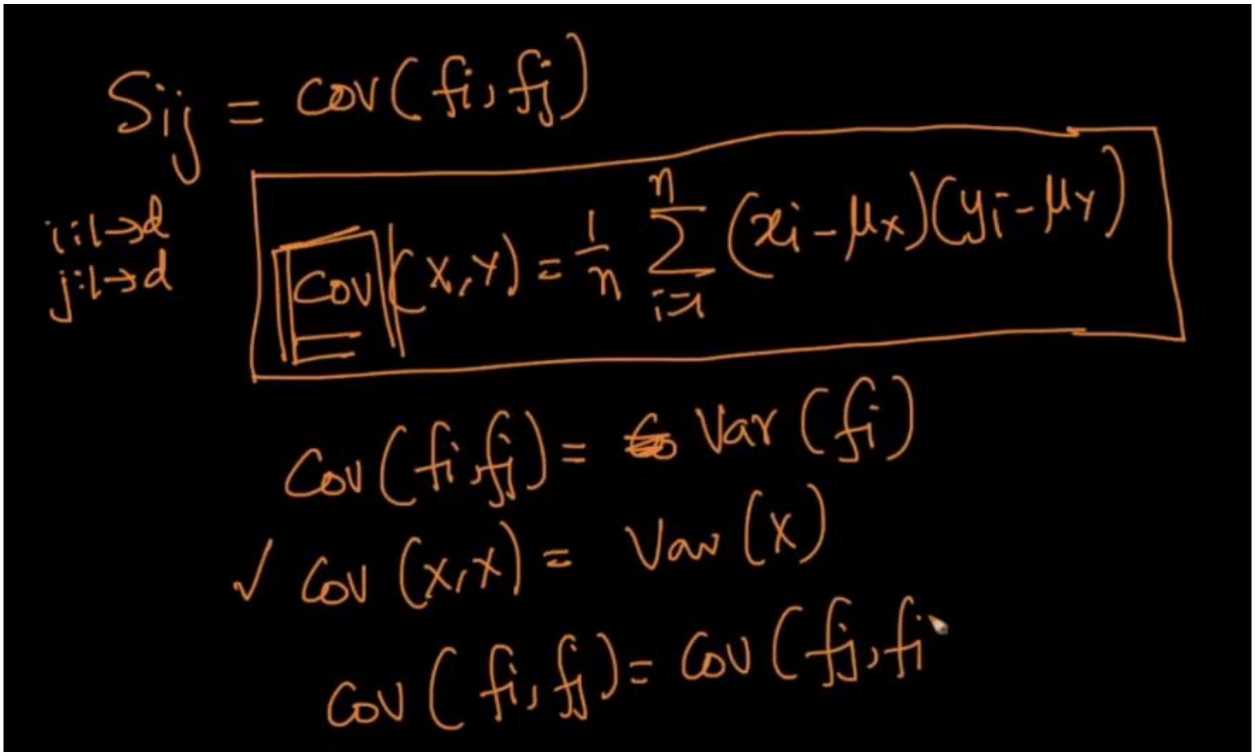
This is one of the interesting data preprocessing in which we find the covariance of each feature with another feature.

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



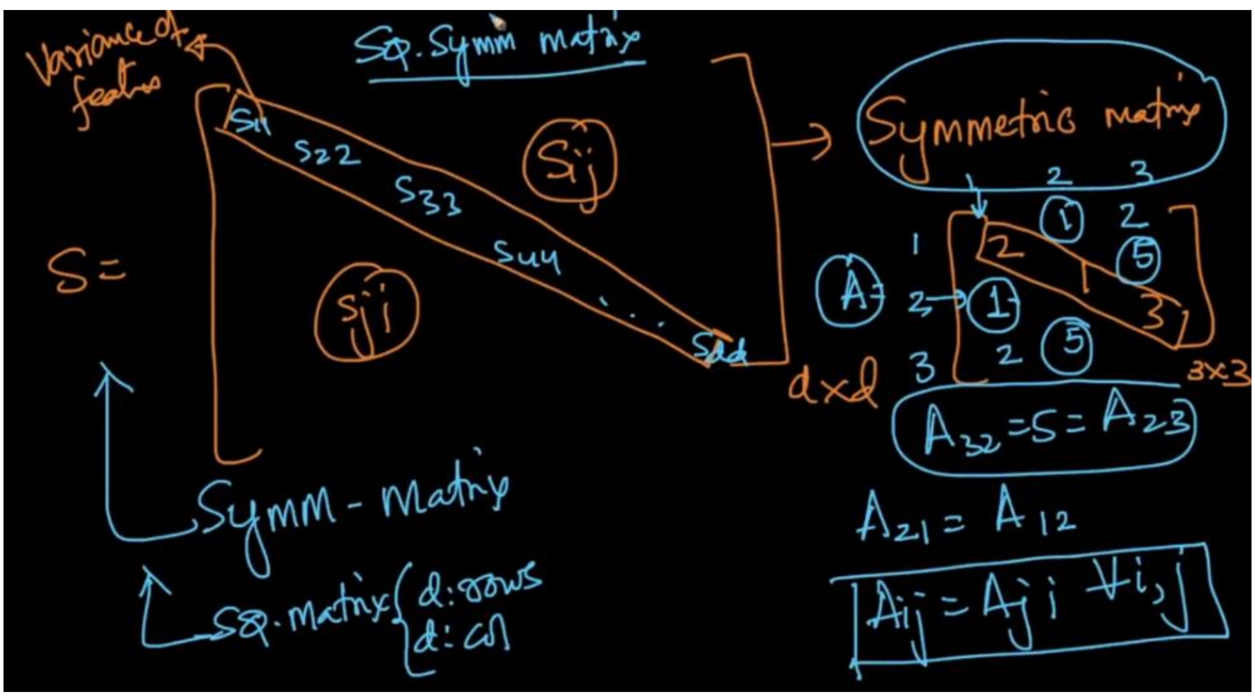
Below figure shows two properties of covariance:

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)



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 metrix $A_{21} = A_{12}$.



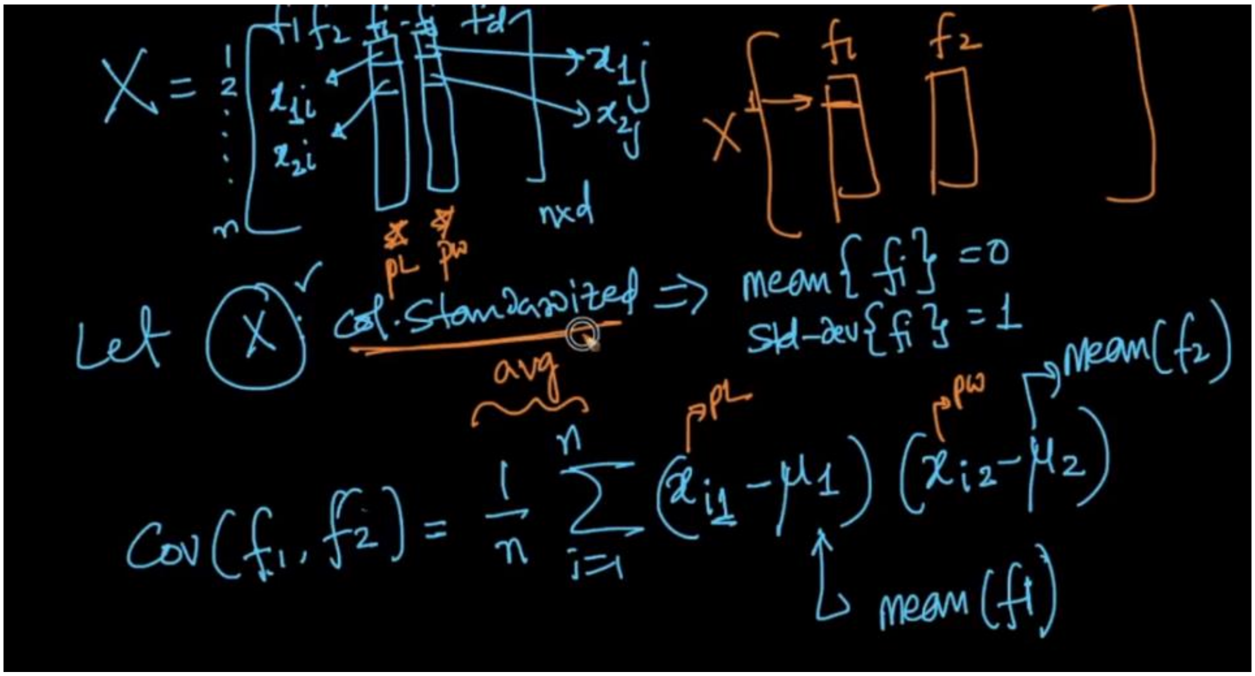
For Finding Co-variance of any data matrix first we'll perform **column standardization**, to bring the mean of all feature to 0.

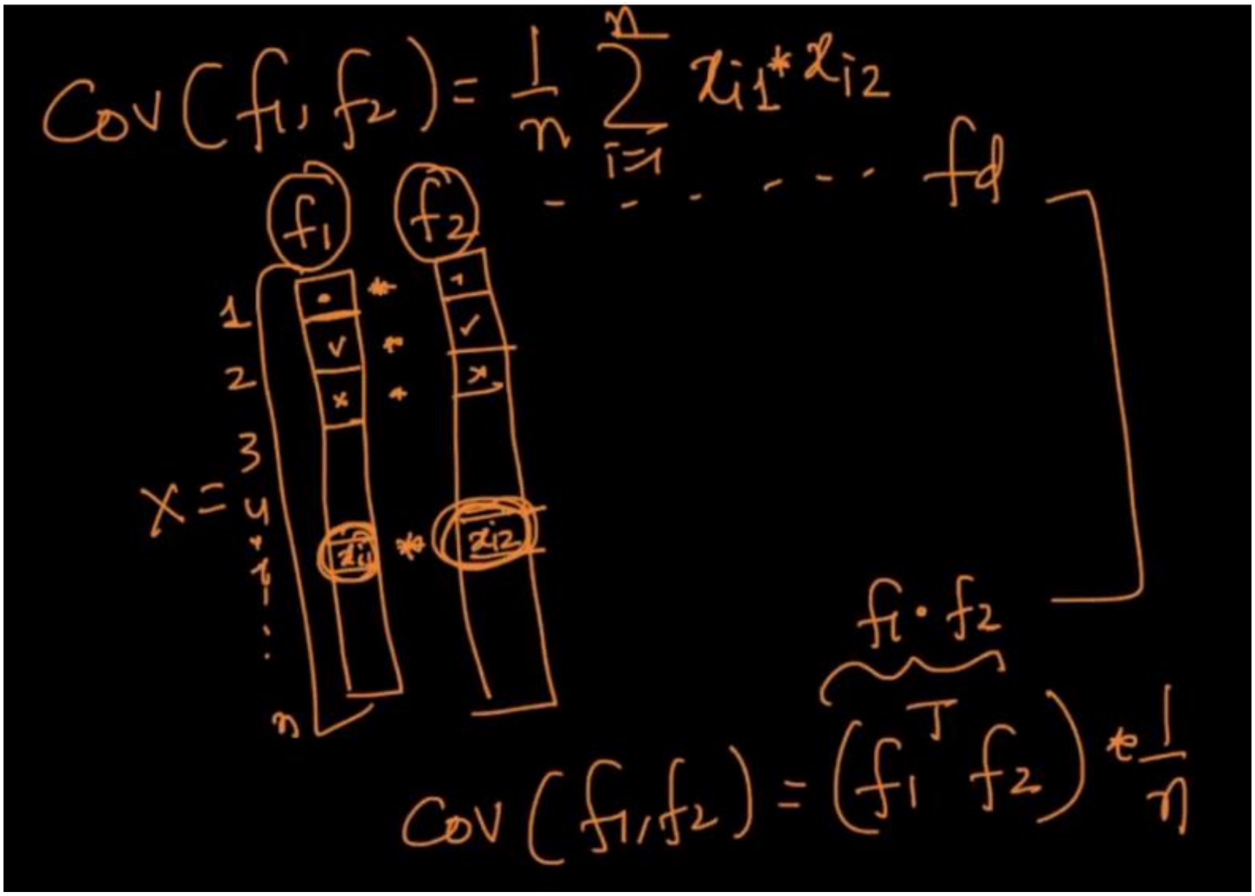
Since after standardization mean becomes 0, so we'll remove mean from co-variance formula,

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

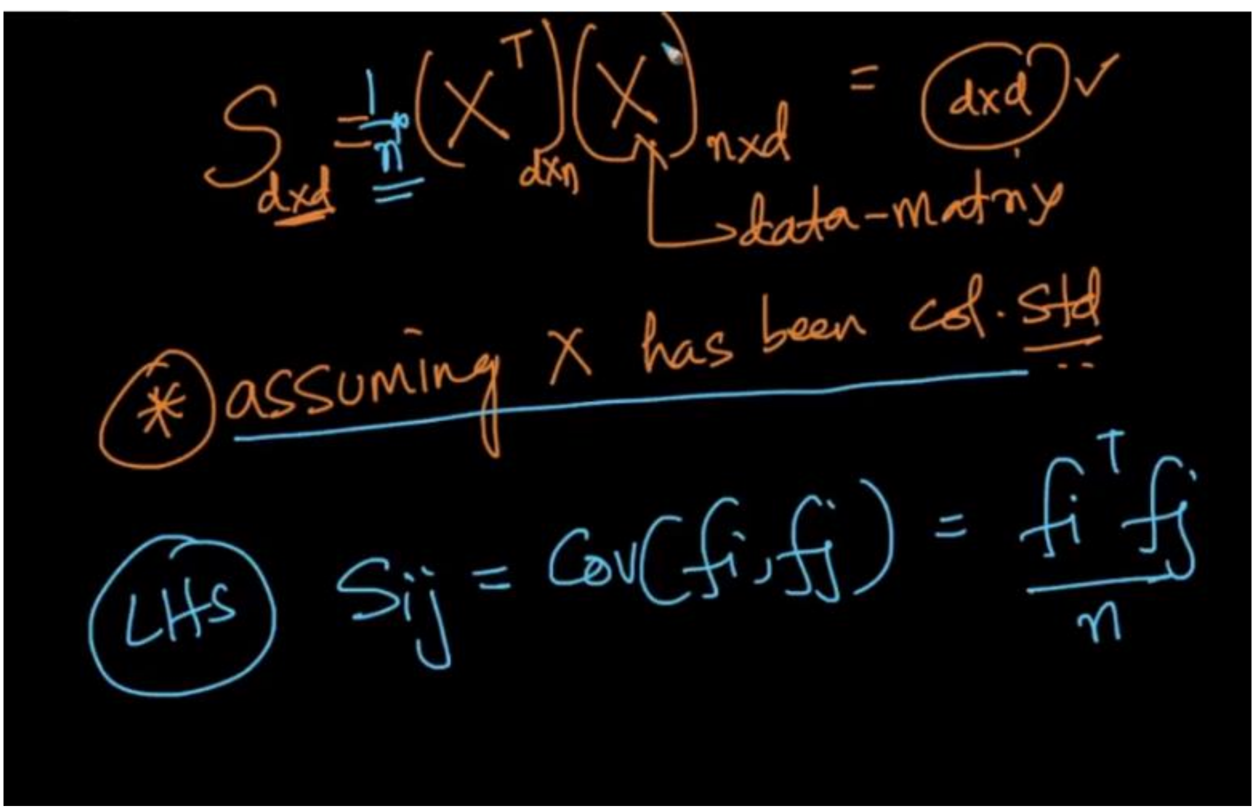
And therefore we can also say that covariance of two features f_1 and f_2 will be dot product as:

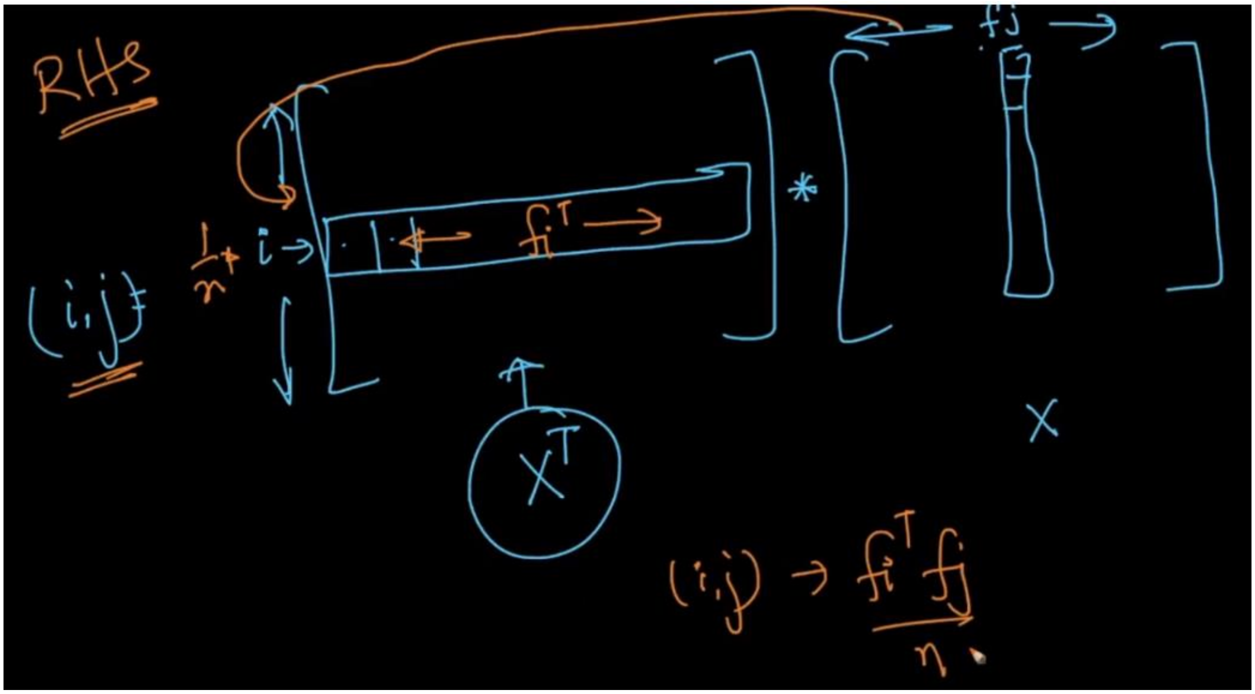
$(f_1 . f_2) / n$

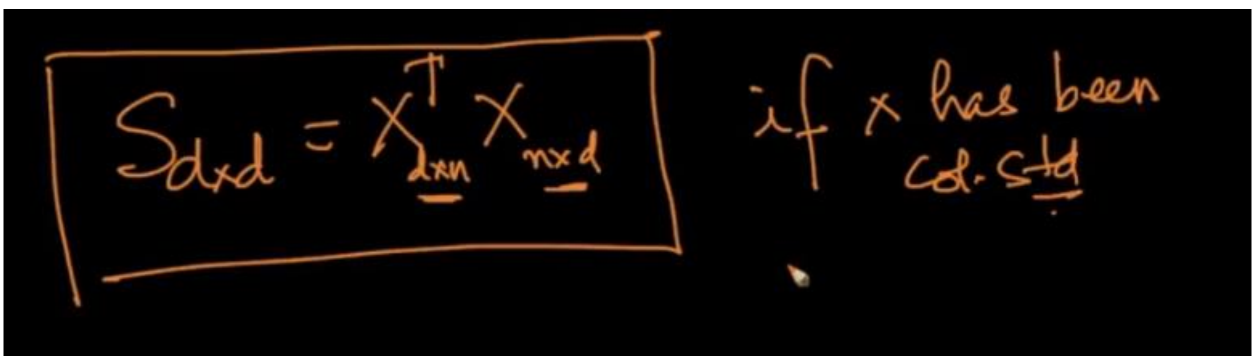




Since after standardization mean becomes 0, and we are making co-varinace matrix just by using value of data-matrix, therefore we can also say tha co-variance matrix can be $(X^T * X)/n$







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.