

Covariance Matrix Computation – Detailed Example

This document explains how to compute the covariance matrix step-by-step using a simple dataset with 3 samples and 2 variables. Covariance measures how much two random variables vary together. The covariance matrix summarizes the variances and covariances among all variables.

Step 1: Dataset

Consider the dataset below with 3 samples (rows) and 2 variables (columns):

Sample	X_1	X_2
1	2	3
2	4	7
3	6	9

Hence, the data matrix X is:

$$X = [[2, 3], [4, 7], [6, 9]]$$

Step 2: Compute the Mean of Each Variable

$$\text{Mean of } X_1 = (2 + 4 + 6) / 3 = 4$$

$$\text{Mean of } X_2 = (3 + 7 + 9) / 3 = 6.333 \text{ (approximately 6.33)}$$

Step 3: Center the Data

Subtract the mean of each variable from its corresponding values to get the centered data matrix.

Sample	$X_1 - 4$	$X_2 - 6.333$
1	-2	-3.333
2	0	0.667
3	2	2.667

Thus, the centered data matrix X_c is:

$$X_c = [[-2, -3.333], [0, 0.667], [2, 2.667]]$$

Step 4: Compute Variances and Covariances

For $n = 3$ samples, we divide by $(n - 1) = 2$.

$$\text{Var}(X_1) = [(-2)^2 + 0^2 + 2^2] / 2 = (4 + 0 + 4) / 2 = 4$$

$$\text{Var}(X_2) = [(-3.333)^2 + (0.667)^2 + (2.667)^2] / 2 = (11.11 + 0.44 + 7.11)/2 = 9.33$$

$$\text{Cov}(X_1, X_2) = [(-2)(-3.333) + (0)(0.667) + (2)(2.667)] / 2 = (6.666 + 0 + 5.334)/2 = 6$$

Step 5: Form the Covariance Matrix

The covariance matrix is symmetric and can be expressed as:

	X_1	X_2
X_1	4	6
X_2	6	9.33

Hence, the covariance matrix is:

$$\text{Cov}(X) = [[4, 6], [6, 9.33]]$$

Step 6: Verification Using Matrix Multiplication

We can also compute the covariance matrix using the matrix formula:

$$\text{Cov}(X) = (1/(n-1)) * X_c^T * X_c$$

where X_c is the centered data matrix.

If we perform the multiplication:

$$\begin{aligned} X_c^T X_c &= [(-2)(-2) + 0^2 + 2^2, (-2)(-3.333) + 0(0.667) + 2(2.667)], \\ &\quad [(-3.333)(-2) + 0.667(0) + 2.667(2), (-3.333)^2 + (0.667)^2 + (2.667)^2] \\ &= [[8, 12], [12, 18.66]] \end{aligned}$$

Dividing by $(n - 1) = 2$, we get:

$$\text{Cov}(X) = [[4, 6], [6, 9.33]]$$

Step 7: Interpretation

The diagonal elements (4 and 9.33) are the variances of X_1 and X_2 respectively. The off-diagonal elements (6) represent the covariance between X_1 and X_2 , indicating a strong positive relationship — as X_1 increases, X_2 tends to increase as well.