Covariance Matrix

A covariance matrix is a type of matrix used to describe the covariance values between two items in a random vector.

The variance-covariance matrix is a square matrix with diagonal elements that represent the variance and the non-diagonal components that express covariance. The covariance of a variable can take any real value- positive, negative, or zero. A positive covariance suggests that the two variables have a positive relationship, whereas a negative covariance indicates that they do not. If two elements do not vary together, they have a zero covariance.

The general form of a covariance matrix is

Covariance Matrix

$$\begin{bmatrix} Var(x_1) & \cdots & Cov(x_n, x_1) \\ \vdots & \ddots & \vdots \\ Cov(x_n, x_1) & \cdots & Var(x_n) \end{bmatrix}$$

Covariance Matrix is a square matrix and is also symmetric in nature

i.e., the transpose of the original matrix gives the original matrix itself.

Covariance Matrix is used in the field of Mathematics, Machine Learning, Finance and Economics. Covariance Matrix is used in Cholskey Decomposition to perform Monte Carlo Simulation which is used to create Mathematical Models.

Examples

Example 1: The data pertains to marks scored by Anna, Caroline, and Laura in Psychology and History. Calculate Covariance Matrix from the data.

Student	Psychology(X)	History(Y)
Anna	80	70
Caroline	63	20
Laura	100	50

Solution:

The following steps have to be followed:

Step 1: Find the mean of variable X.

Sum up all the observations in variable X and divide the sum obtained with the number of terms. Thus, (80 + 63 + 100)/3 = 81.

Step 2: Subtract the mean from all observations.

(80 – 81), (63 – 81), (100 – 81).

Step 3: Take the squares of the differences obtained above and then add them up.

Thus, $(80 - 81)^2 + (63 - 81)^2 + (100 - 81)^2$.

Step 4: Find the variance of X by dividing the value obtained in Step 3 by 1 less than the total number of observations.

$var(X) = [(80 - 81)^2 + (63 - 81)^2 + (100 - 81)^2] / (3 - 1) = 343$.

Step 5: Similarly, repeat steps 1 to 4 to calculate the variance of Y.

Var(Y) = 633.

Step 6: Choose a pair of variables.

Step 7: Subtract the mean of the first variable (X) from all observations;

(80 – 81), (63 – 81), (100 – 81).

Step 8: Repeat the same for variable Y; (70 – 47), (20 – 47), (50 – 47).

Step 9: Multiply the corresponding terms:

(80 – 81)(70 – 47), (63 – 81)(20 – 47), (100 – 81)(50 – 47).

Step 10: Find the covariance by adding these values and dividing them by (n – 1).

Cov(X, Y) = (80 – 81)(70 – 47) + (63 – 81)(20 – 47) + (100 – 81)(50 – 47)/3-1

= 481.

Step 11: Use the general formula for the covariance matrix to arrange the terms.

covariance matrix = $\begin{bmatrix} 343 & 481 \\ 481 & 633 \end{bmatrix}$

Example 2: The marks scored by 3 students in Physics and Biology are given below:

Student	Physics(X)	Biology(Y)
A	92	80
B	60	30
C	100	70

Calculate Covariance Matrix from the above data.

Solution:

Here, $\mu_x = 84$, n = 3

var(x) = $[(92 - 84)^2 + (60 - 84)^2 + (100 - 84)^2] / (3 - 1) = 448$

Also, $\mu_y = 60$, n = 3

var(y) = $[(80 - 60)^2 + (30 - 60)^2 + (70 - 60)^2] / (3 - 1) = 700$

cov(x, y) = $[(92 - 84)(80 - 60) + (60 - 84)(30 - 60) + (100 - 84)(70 - 60)] / (3 - 1)$

= 520.

covariance matrix = $\begin{bmatrix} 448 & 520 \\ 520 & 700 \end{bmatrix}$

Example 3. Interpret the following covariance matrix:

$$\begin{array}{c c} & \begin{array}{ccc} X & Y & Z \end{array} \\ \begin{array}{c} X \\ Y \\ Z \end{array} & \begin{bmatrix} 60 & 32 & -4 \\ 32 & 30 & 0 \\ -4 & 0 & 80 \end{bmatrix} \end{array}$$

Solution:

1. The diagonal elements 60, 30, and 80 indicate the variance in data sets X, Y, and Z respectively. Y shows the lowest variance whereas Z displays the highest variance.
2. The covariance for X and Y is 32. As this is a positive number it means that when X increases (or decreases) Y also increases (or decreases)

3. The covariance for X and Z is –4. As it is a negative number it implies that when X increases Z decreases and vice-versa.
4. The covariance for Y and Z is 0. This means that there is no predictable relationship between the two data sets.

Python code to demonstrate the use of numpy.cov

```
import numpy as np

x = [1.23, 2.12, 3.34, 4.5]

y = [2.56, 2.89, 3.76, 3.95]

# Find out covariance with respect columns

cov_mat = np.stack((x, y), axis = 0)

print(np.cov(cov_mat))
```

Output:

```
[[ 2.03629167  0.9313]
 [ 0.9313      0.4498]]
```

Python code to demonstrate the use of numpy.cov

```
import numpy as np

x = [1.23, 2.12, 3.34, 4.5]

y = [2.56, 2.89, 3.76, 3.95]

# Find out covariance with respect rows

cov_mat = np.stack((x, y), axis = 1)

print("shape of matrix x and y:", np.shape(cov_mat))

print("shape of covariance matrix:", np.shape(np.cov(cov_mat)))

print(np.cov(cov_mat))
```

Output:

```
shape of matrix x and y: (4, 2)

shape of covariance matrix: (4, 4)

[[ 0.88445  0.51205  0.2793   -0.36575]

 [ 0.51205  0.29645  0.1617   -0.21175]

 [ 0.2793   0.1617   0.0882   -0.1155]

 [-0.36575 -0.21175 -0.1155    0.15125]]
```

Example

The Following table lists the weight and heights of 5 boys. Find the covariance matrix for the data.

Boy	1	2	3	4	5
Weight(lb)	120	125	125	135	145
Height(in.)	61	60	64	68	72

```
import numpy as np

# Data

x = [120,125,125,135,145]

y = [61,60,64,68,72]
```

Create a 2D array

X = np.stack((x, y), axis = 0)

Calculate the covariance matrix

cov_matrix = np.cov(X)

print("Covariance matrix:\n", cov_matrix)

output is

```
Covariance matrix:
 [[100.    47.5]
 [ 47.5   25. ]]
```