Principal component analysis is a dimensionality reduction technique that enables you to identify correlations and patterns in a dataset so it can be transformed into a dataset of significantly lower dimensions without any loss of any important information.

Step by Step process:

1. Take the whole dataset consisting of input variables to be used to train the model. Drop target variable.

2. Compute the mean for every dimension of the whole dataset.

3. Compute the covariance matrix of the whole dataset.

4. Compute eigenvectors and the corresponding eigenvalues.

5. Sort the eigenvectors by decreasing eigenvalues and choose k eigenvectors with the largest eigenvalues to form a d × k dimensional matrix W.

Use this d × k eigenvector matrix to transform the samples onto the new subspace.

Mean of matrix A

Column X = = 80

Column Y = = 70

Column Z = = 60

Covariance of a matrix formula



|  |  |  |
| --- | --- | --- |
| X - mean of column X | Y – mean of column Y | Z – mean of column Z |
| 90-80 = 10 | 60-70 = -10 | 90-60 = 30 |
| 90-80 = 10 | 90-70 = 20 | 30-60 = -30 |
| 60-80 = -20 | 60-70 =-10 | 60 – 60 = 0 |

Cov(X,Y) = = 150

Cov(X,Z) = = 0

Cov(Y,Z) = = 150

Cov(X,X) = = 300

Cov(Y,Y) = = 300

Cov(Z,Z) = =900

Hence, Covariance of matrix:

Calculating the eigenvalues and eigenvectors

det(**A-I *λ****) = 0*

Where A = covariance of matrix

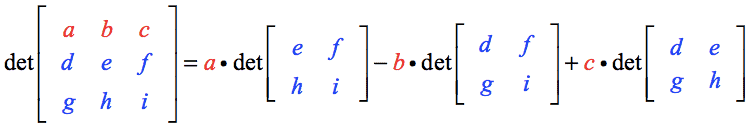
**I =** Identity matrix

***Λ* = Eigenvalue**

- ***λ***

-

The determinant of a matrix:



300-\* – 150\* + 0\*

300-\* - 150\* +0\*

300-\* - 150 + 0

300-\* - 20,250,000 + 22,500

– 20,250,000 + 22,500

81,000,000 – 630,000 - 150\*\*2 - – 20,250,000 + 22,500

60,750,000 – 607,500 - 1500 - = 0

Rewriting the equation

-

Solving this equation gives the following eigenvalues:

Solving for the eigenvectors(X,Y,Z)

When

=

=

At X = 1:

450 \*1 + (150\*Y) + (0\*Z) = 0 ----eqn1

150 + 450Y +150Z = 0 --------eqn 2

Solving for Y

450 + 150Y = 0

150Y = -450

Y = -450/150

Y = -3

Put Y = -3 in eqn 2

150+(450\*-3)+150Z = 0

150 + 1350 +150Z = 0

150Z =- 1500

Z = -1500/150

Z= -10

Eigenvector at :

When

=

=

At X = 1:

750 \*1 + (150\*Y) + (0\*Z) = 0 ----eqn1

150 + 750Y +150Z = 0 --------eqn 2

Solving for Y

750 + 150Y = 0

150Y = -750

Y = -750/150

Y = -5

Put Y = -5 in eqn 2

150+(750\*-5)+150Z = 0

150 - 3750 +150Z = 0

150Z = 3600

Z = 3600/150

Z= 24

Eigenvector at :

When

=

=

At X = 1:

1200 \*1 + (150\*Y) + (0\*Z) = 0 ----eqn1

150 + 1200Y +150Z = 0 --------eqn 2

Solving for Y

1200 + 150Y = 0

150Y = -1200

Y = -1200/150

Y = -8

Put Y = -8 in eqn 2

150+(1200\*-8)+150Z = 0

150 - 9600 +150Z = 0

150Z = -9450

Z = -9450/150

Z= 63

Eigenvector at :

Principal Components: Sort eigenvalues in descending order

Hence, principal components become: