[Principal Component Analysis Numerical Example | Gate Vidyalay](https://www.gatevidyalay.com/tag/principal-component-analysis-numerical-example/)

**Example-**

Consider the following example-

* The following graph shows two dimensions x1 and x2.
* x1 represents the measurement of several objects in cm.
* x2 represents the measurement of several objects in inches.

A picture containing chart

Description automatically generated

In machine learning,

* Using both these dimensions convey similar information.
* Also, they introduce a lot of noise in the system.
* So, it is better to use just one dimension.

Using dimension reduction techniques-

* We convert the dimensions of data from 2 dimensions (x1 and x2) to 1 dimension (z1).
* It makes the data relatively easier to explain.

Logo

Description automatically generated with medium confidence

**Benefits-**

Dimension reduction offers several benefits such as-

* It compresses the data and thus reduces the storage space requirements.
* It reduces the time required for computation since less dimensions require less computation.
* It eliminates the redundant features.
* It improves the model performance.

**Dimension Reduction Techniques-**

The two popular and well-known dimension reduction techniques are-

Text

Description automatically generated

1. Principal Component Analysis (PCA)
2. Fisher Linear Discriminant Analysis (LDA)

In this article, we will discuss about Principal Component Analysis.

**Principal Component Analysis-**

* Principal Component Analysis is a well-known dimension reduction technique.
* It transforms the variables into a new set of variables called as principal components.
* These principal components are linear combination of original variables and are orthogonal.
* The first principal component accounts for most of the possible variation of original data.
* The second principal component does its best to capture the variance in the data.
* There can be only two principal components for a two-dimensional data set.

**PCA Algorithm-**

The steps involved in PCA Algorithm are as follows-

**Step-01:** Get data.

**Step-02:** Compute the mean vector (µ).

**Step-03:** Subtract mean from the given data.

**Step-04:** Calculate the covariance matrix.

**Step-05:** Calculate the eigen vectors and eigen values of the covariance matrix.

**Step-06:** Choosing components and forming a feature vector.

**Step-07:** Deriving the new data set.

**PRACTICE PROBLEMS BASED ON PRINCIPAL COMPONENT ANALYSIS-**

**Problem-01:**

Given data = { 2, 3, 4, 5, 6, 7 ; 1, 5, 3, 6, 7, 8 }.

Compute the principal component using PCA Algorithm.

**OR**

Consider the two dimensional patterns (2, 1), (3, 5), (4, 3), (5, 6), (6, 7), (7, 8).

Compute the principal component using PCA Algorithm.

**OR**

Compute the principal component of following data-

CLASS 1

X = 2 , 3 , 4

Y = 1 , 5 , 3

CLASS 2

X = 5 , 6 , 7

Y = 6 , 7 , 8

**Solution-**

We use the above discussed PCA Algorithm-

**Step-01:**

Get data.

The given feature vectors are-

* x1 = (2, 1)
* x2 = (3, 5)
* x3 = (4, 3)
* x4 = (5, 6)
* x5 = (6, 7)
* x6 = (7, 8)

A picture containing text, device, gauge, meter

Description automatically generated

**Step-02:**

Calculate the mean vector (µ).

Mean vector (µ)

= ((2 + 3 + 4 + 5 + 6 + 7) / 6, (1 + 5 + 3 + 6 + 7 + 8) / 6)

= (4.5, 5)

Thus,

A picture containing text, device, gauge

Description automatically generated

**Step-03:**

Subtract mean vector (µ) from the given feature vectors.

* x1 – µ = (2 – 4.5, 1 – 5) = (-2.5, -4)
* x2 – µ = (3 – 4.5, 5 – 5) = (-1.5, 0)
* x3 – µ = (4 – 4.5, 3 – 5) = (-0.5, -2)
* x4 – µ = (5 – 4.5, 6 – 5) = (0.5, 1)
* x5 – µ = (6 – 4.5, 7 – 5) = (1.5, 2)
* x6 – µ = (7 – 4.5, 8 – 5) = (2.5, 3)

Feature vectors (xi) after subtracting mean vector (µ) are-

A picture containing text, gauge, device, meter

Description automatically generated

**Step-04:**

Calculate the covariance matrix.

Covariance matrix is given by-

A picture containing logo

Description automatically generated

Now,

A picture containing diagram

Description automatically generated

Now,

Covariance matrix

= (m1 + m2 + m3 + m4 + m5 + m6) / 6

On adding the above matrices and dividing by 6, we get-

A picture containing text, device

Description automatically generated

**Step-05:**

Calculate the eigen values and eigen vectors of the covariance matrix.

λ is an eigen value for a matrix M if it is a solution of the characteristic equation |M – λI| = 0.

So, we have-

A picture containing text, device, meter

Description automatically generated

From here,

(2.92 – λ)(5.67 – λ) – (3.67 x 3.67) = 0

16.56 – 2.92λ – 5.67λ + λ2 – 13.47 = 0

λ2 – 8.59λ + 3.09 = 0

Solving this quadratic equation, we get λ = 8.22, 0.38

Thus, two eigen values are λ1 = 8.22 and λ2 = 0.38.

Clearly, the second eigen value is very small compared to the first eigen value.

So, the second eigen vector can be left out.

Eigen vector corresponding to the greatest eigen value is the principal component for the given data set.

So. we find the eigen vector corresponding to eigen value λ1.

We use the following equation to find the eigen vector-

MX = λX

where-

* M = Covariance Matrix
* X = Eigen vector
* λ = Eigen value

Substituting the values in the above equation, we get-

A picture containing text, sign, device, gauge

Description automatically generated

Solving these, we get-

2.92X1 + 3.67X2 = 8.22X1

3.67X1 + 5.67X2 = 8.22X2

On simplification, we get-

5.3X1 = 3.67X2 ………(1)

3.67X1 = 2.55X2 ………(2)

From (1) and (2), **X1 = 0.69X2**

From (2), the eigen vector is-

A picture containing text, sign, gauge

Description automatically generated

Thus, principal component for the given data set is-

A picture containing text, sign, device, gauge

Description automatically generated

Lastly, we project the data points onto the new subspace as-

A screenshot of a computer

Description automatically generated with medium confidence