Principal Component Analysis

Dr. Avinash Kumar Singh

Founder, Robotics and Artificial Intelligence Training Academy

Senior Researcher Montpellier University France

Discussion Points

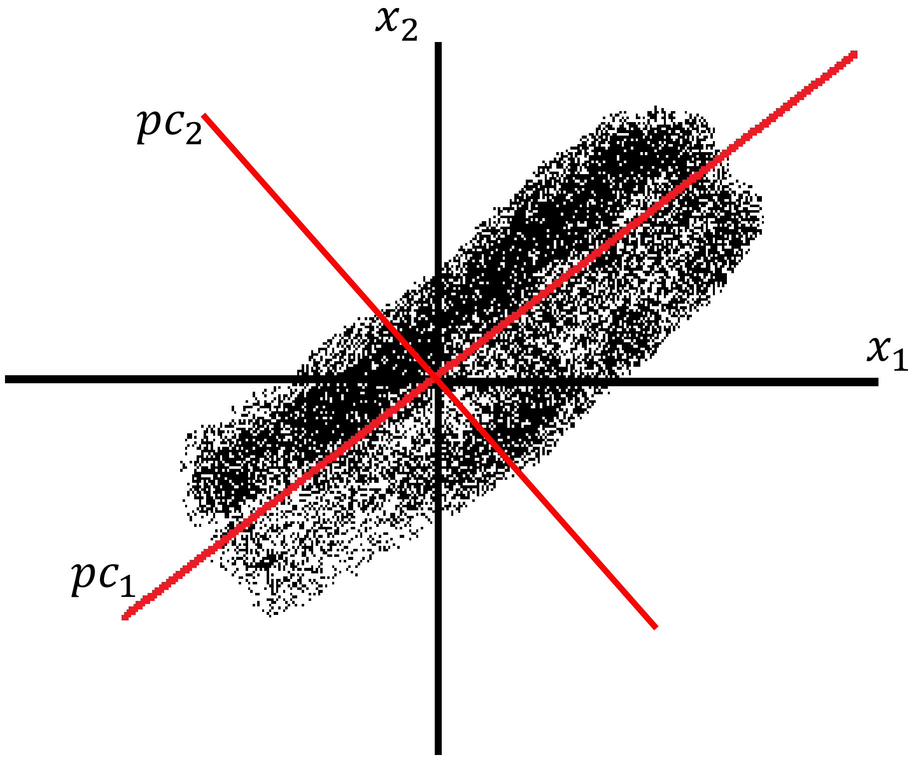
- Machine Learning
    - What, Why and Types
    - Mathematics
    - Applications

Objective

- Used to reduce the dimension of multivariate data, while preserving as much of the relevant information as possible.
  - If we have samples represented in the m dimensional space, $\{x_1, x_2, x_3, x_4, x_5, x_6, \ldots x_M\}$, our intension is to find out k features, such that $k \leq m$ that preserve most of the variance present in the data.
- Good tool for feature extraction.

Summary

- We should select those directions where variance is maximum.
- Each direction will give one principal component.
- First principal component has maximum variance, Second has 2nd maximum variance, and orthogonal to first, Third has 3rd maximum variance and orthogonal to first and second one, like this we will select k component.



Derivation

- In many physical, biological, and statistical convention it is desirable to represent a system of points with the help of a line or plane.
- Therefore, we can represent the equation of line with the linear combination of these points(variables).

$y_k = \sum_{i=1}^{m} ak_i x_i$ Where K=1,2,3, ……. p

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ . \\ y_p \end{bmatrix} = \begin{bmatrix} a1_1 & a1_2 & a1_3 & . & a1_m \\ a2_1 & a2_2 & a2_3 & . & a2_m \\ a3_1 & a3_2 & a3_3 & . & a3_m \\ . & . & . & . & . \\ ap_1 & ap_2 & ap_3 & . & ap_m \end{bmatrix} * \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ . \\ x_m \end{bmatrix}$$

Steps used in PCA

- Mean of the data (along each feature):

Let us assume, we have samples having m features. We have stored all these in a matrix called training samples T having N rows and M columns (where N represents the population and M represents the Features).So in that regard we have data like.

$a1_1$	$a1_2$	$a1_3$	.	$a1_m$
$a2_1$	$a2_2$	$a2_3$	.	$a2_m$
$a3_1$	$a3_2$	$a3_3$	.	$a3_m$
.	.	.	.	.
ap_1	ap_2	ap_3	.	ap_m
μ_1	μ_2	μ_3	μ_{m-1}	μ_m

$$(\mu)_{1*m} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{\mathrm{p}} \mathrm{T}(i,j)}{p}$$

Here

: i represents features

: j represents observation.

: $T(i,j)$ represents the value of i^{th} feature at j^{th} observation

Step-1: Mean Aligned Data

- Subtract the from each sample to do the mean zero of the data.

$a1_1$	$a1_2$	$a1_3$	.	$a1_m$
$a2_1$	$a2_2$	$a2_3$	.	$a2_m$
$a3_1$	$a3_2$	$a3_3$	.	$a3_m$
.	.	.	.	.
ap_1	ap_2	ap_3	.	ap_m

$-$

μ_1	μ_2	μ_3	μ_{m-1}	μ_m

$=$

$\boxed{\text{D}}$

$$\begin{array}{|c|}\hline d_{11} \\ \hline d_{12} \\ \hline d_{13} \\ \hline . \\ \hline d_{1m} \\ \hline \end{array} = \begin{array}{|c|}\hline a1_1 \\ \hline a1_2 \\ \hline a1_3 \\ \hline . \\ \hline a1_m \\ \hline \end{array} - \begin{array}{|c|}\hline \mu_1 \\ \hline \mu_2 \\ \hline \mu_3 \\ \hline . \\ \hline \mu_m \\ \hline \end{array}$$

$\boxed{\text{D}} =$

d_{11}	d_{12}	d_{13}	.	d_{1m}
d_{21}	d_{22}	d_{23}	.	d_{2m}
d_{31}	d_{32}	d_{33}	.	d_{3m}
d_{41}	d_{42}	d_{43}	.	d_{4m}
d_{p1}	d_{p2}	d_{p3}	.	d_{pm}

Step-2: Calculate Variance, Covariance

- Variance (σ) denotes how the data is distributed along its mean, while Co-variance (Σ) shows how the data is relate to other.

$$\underbrace{\begin{bmatrix} d_{11} \\ d_{21} \\ d_{31} \\ d_{41} \\ d_{p1} \end{bmatrix} = \begin{bmatrix} a1_{1} \\ a2_{1} \\ a3_{1} \\ . \\ ap_{1} \end{bmatrix} - \begin{bmatrix} \mu_1 \\ \mu_1 \\ \mu_1 \\ \mu_1 \\ \mu_1 \end{bmatrix}}_{\text{Mean Alignment}}$$

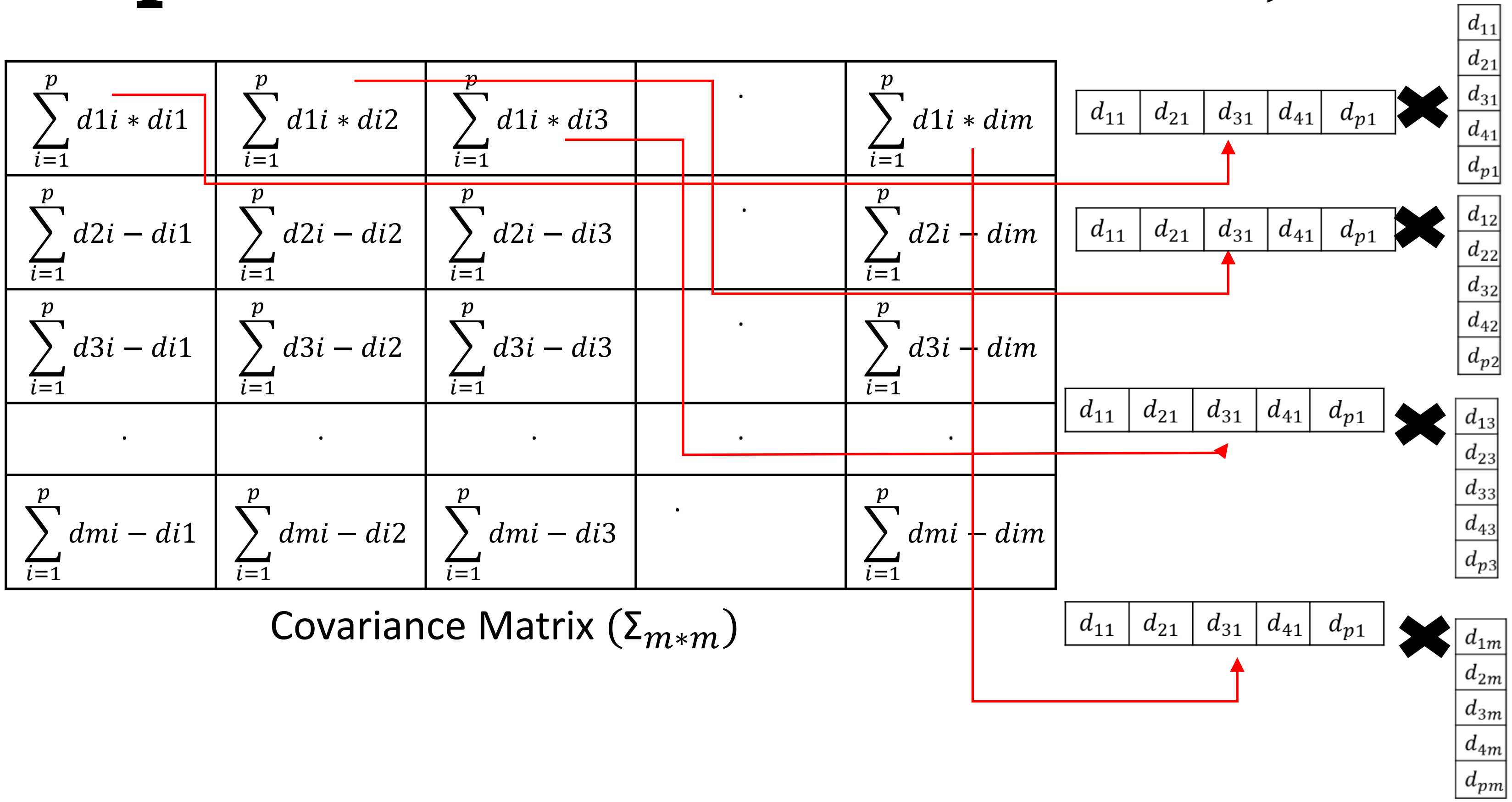
→

$$d_{11} * d_{11} + \overbrace{d_{12} * d_{21} + d_{13} * d_{31} + d_{14}}^{\text{Variance }(\sigma)=\sum_{i=1}^{p} d1i * di1} * d_{41} + d_{1p} * d_{p1}$$

$$=$$

$$\underbrace{\begin{bmatrix} d_{11} & d_{12} & d_{13} & d_{14} & d_{1p} \end{bmatrix}}_{\boxed{D}^{T}} \times \underbrace{\begin{bmatrix} d_{11} \\ d_{21} \\ d_{31} \\ d_{41} \\ d_{p1} \end{bmatrix}}_{\boxed{D}}$$

Step-2: Calculate Variance, Covariance



$\sum_{i=1}^{p} d1i * di1$	$\sum_{i=1}^{p} d1i * di2$	$\sum_{i=1}^{p} d1i * di3$	.	$\sum_{i=1}^{p} d1i * dim$
$\sum_{i=1}^{p} d2i - di1$	$\sum_{i=1}^{p} d2i - di2$	$\sum_{i=1}^{p} d2i - di3$	.	$\sum_{i=1}^{p} d2i - dim$
$\sum_{i=1}^{p} d3i - di1$	$\sum_{i=1}^{p} d3i - di2$	$\sum_{i=1}^{p} d3i - di3$	.	$\sum_{i=1}^{p} d3i - dim$
.	.	.	.	.
$\sum_{i=1}^{p} dmi - di1$	$\sum_{i=1}^{p} dmi - di2$	$\sum_{i=1}^{p} dmi - di3$	.	$\sum_{i=1}^{p} dmi - dim$

Covariance Matrix (Σ_{m*m})

$$\begin{bmatrix} d_{11} & d_{21} & d_{31} & d_{41} & d_{p1} \end{bmatrix} \times \begin{bmatrix} d_{11} \\ d_{21} \\ d_{31} \\ d_{41} \\ d_{p1} \end{bmatrix}$$

$$\begin{bmatrix} d_{11} & d_{21} & d_{31} & d_{41} & d_{p1} \end{bmatrix} \times \begin{bmatrix} d_{12} \\ d_{22} \\ d_{32} \\ d_{42} \\ d_{p2} \end{bmatrix}$$

$$\begin{bmatrix} d_{11} & d_{21} & d_{31} & d_{41} & d_{p1} \end{bmatrix} \times \begin{bmatrix} d_{13} \\ d_{23} \\ d_{33} \\ d_{43} \\ d_{p3} \end{bmatrix}$$

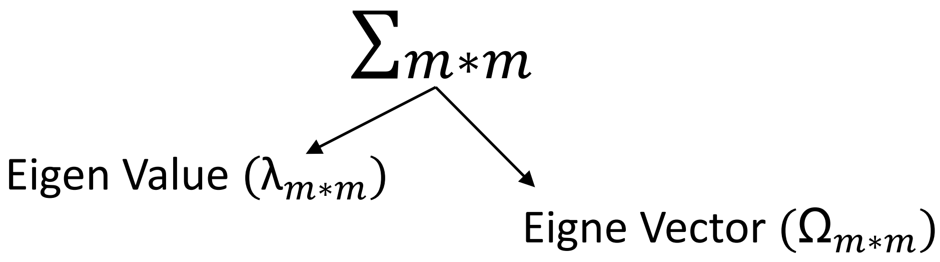
$$\begin{bmatrix} d_{11} & d_{21} & d_{31} & d_{41} & d_{p1} \end{bmatrix} \times \begin{bmatrix} d_{1m} \\ d_{2m} \\ d_{3m} \\ d_{4m} \\ d_{pm} \end{bmatrix}$$

d_{11}	d_{12}	d_{13}	.	d_{1m}
d_{21}	d_{22}	d_{23}	.	d_{2m}
d_{31}	d_{32}	d_{33}	.	d_{3m}
d_{41}	d_{42}	d_{43}	.	d_{4m}
d_{p1}	d_{p2}	d_{p3}	.	d_{pm}

Lookup table

Step-3: Eigen Vector and Eigen Value Decomposition

- From the above equation we can say that co-variance can be calculated as $(\overline{Y} \times \overline{Y}^{T})$, resultant $(m * m)$ dimension. Let's say the covariance matrix is $\sum$
- Principal components are the Eigen values and Eigen vectors, those are computed on the basis of co-variance matrix, calculated in the previous step.



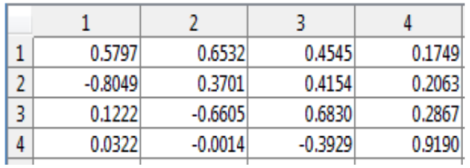
$\sum_{m*m}$

Eigen Value (λ_{m*m})

Eigne Vector (Ω_{m*m})

Example of Eigen value and vector

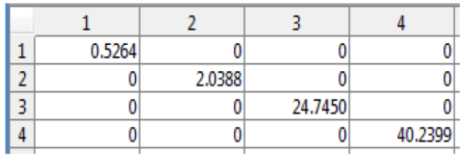
Eigen Vector →



	1	2	3	4
1	0.5797	0.6532	0.4545	0.1749
2	-0.8049	0.3701	0.4154	0.2063
3	0.1222	-0.6605	0.6830	0.2867
4	0.0322	-0.0014	-0.3929	0.9190

Eigen Values →

Diagonal elements



	1	2	3	4
1	0.5264	0	0	0
2	0	2.0388	0	0
3	0	0	24.7450	0
4	0	0	0	40.2399

Step:4 Eigen Vector Selection

- Sort the eigen values in the descending order and the eigen vector as well.
- We should select those λ's which have maximum values, because this shows the variance. Hence for selecting the best principal components (k), we should define a threshold above which we can select all principal components such that (λ>=TH).
- What should be the optimal value for selecting k ?

$$\boldsymbol{Variance} = \sum \lambda_{\boldsymbol{i}}$$

- Combination of these eigenvectors are known as feature vectors, say "W"

Step 5: Projecting the data to principal directions

$$W=\begin{array}{|c|c|c|c|c|}\hline w1_1 & w1_2 & w1_3 & & w1_m \\ \hline w2_1 & w2_2 & w2_3 & & w2_m \\ \hline w3_1 & w3_2 & w3_3 & & w3_m \\ \hline . & . & . & & . \\ \hline wm_1 & wm_2 & wm_3 & & wm_m \\ \hline \end{array}$$

m*m

$$W_k=\begin{array}{|c|c|c|}\hline w1_1 & w1_2 & w1_3 \\ \hline w2_1 & w2_2 & w2_3 \\ \hline w3_1 & w3_2 & w3_3 \\ \hline . & . & . \\ \hline wm_1 & wm_2 & wm_3 \\ \hline \end{array}$$

m*k

$$ProjectedSamples_{p*k} = D_{p*m} * W_{m*k}$$

$$\begin{array}{|c|c|c|c|c|}\hline d_{11} & d_{12} & d_{13} & . & d_{1m} \\ \hline d_{21} & d_{22} & d_{23} & . & d_{2m} \\ \hline d_{31} & d_{32} & d_{33} & . & d_{3m} \\ \hline d_{41} & d_{42} & d_{43} & . & d_{4m} \\ \hline d_{p1} & d_{p2} & d_{p3} & . & d_{pm} \\ \hline \end{array}$$

$$\times$$

$$\begin{array}{|c|c|c|}\hline w1_1 & w1_2 & w1_3 \\ \hline w2_1 & w2_2 & w2_3 \\ \hline w3_1 & w3_2 & w3_3 \\ \hline . & . & . \\ \hline wm_1 & wm_2 & wm_3 \\ \hline \end{array}$$

Constraint

- Linearity:

We can apply PCA to classify the data which is linearly separable, to deal with non-linear data we can use karnel PCA.

- Only suits for the Gaussian distribution:

In PCA we took assumption like mean should be zero and maximum variance will be 1, this assumption only holds when the distribution followed by the data is Gaussian.