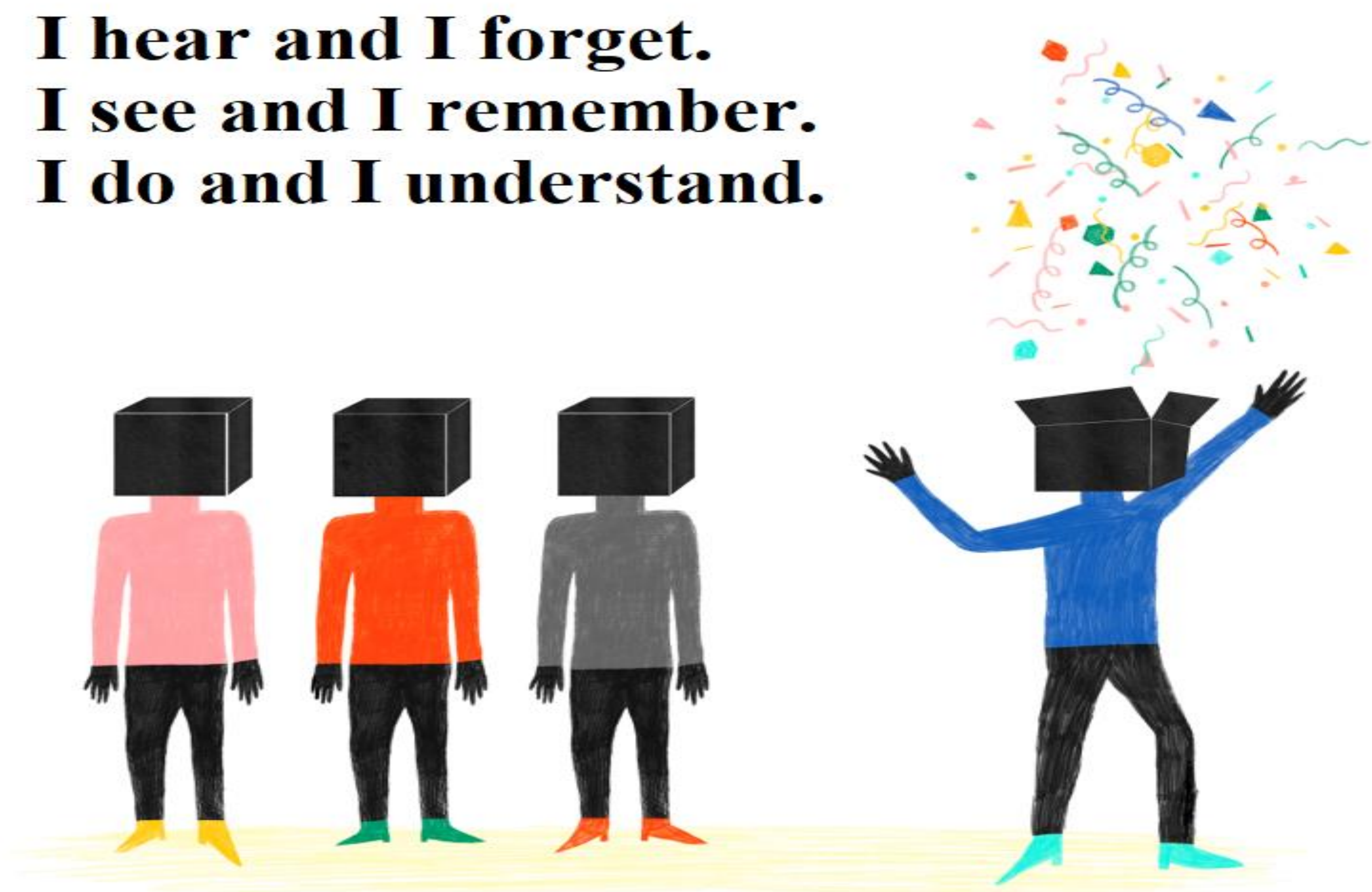
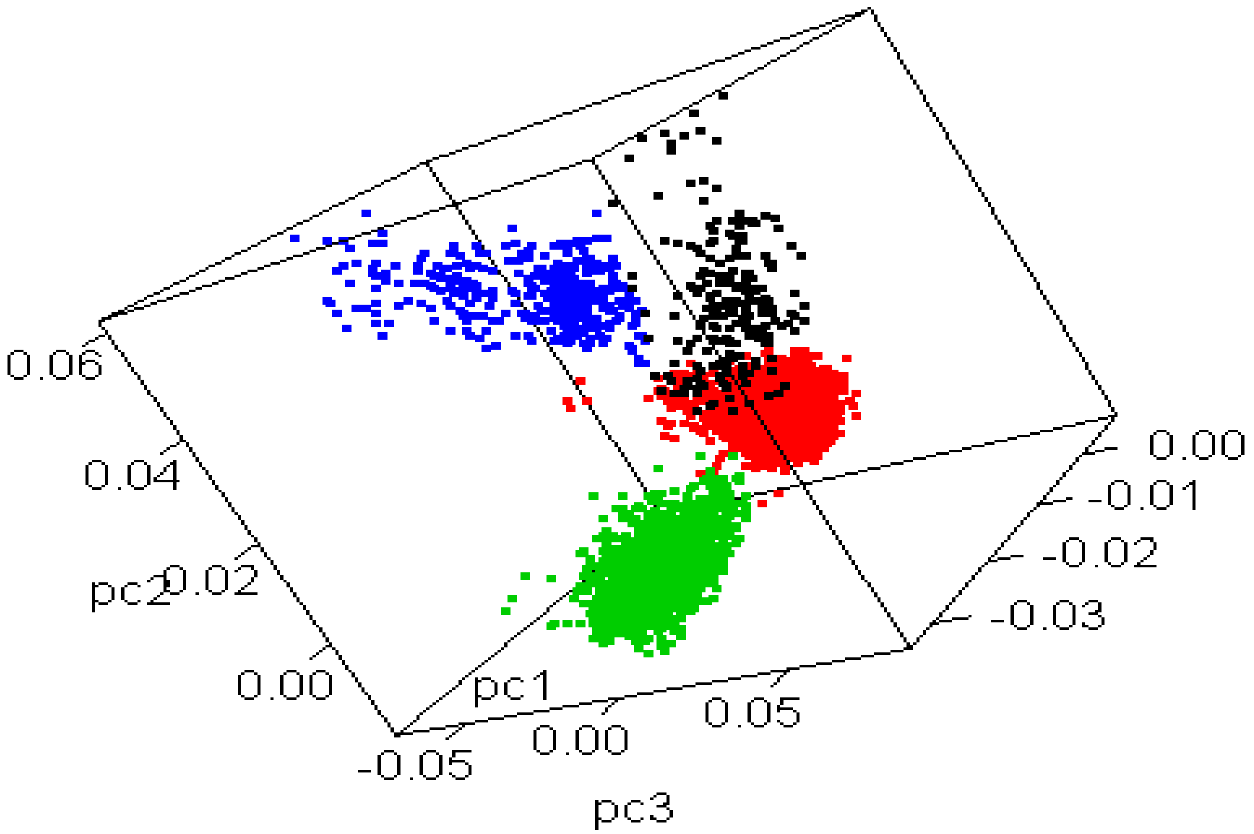


Chandan Verma

Corporate Trainer(Machine Learning,AI,Cloud Computing,IOT)

www.facebook.com/verma.chandan.070

Dimension Reduction



Problem of High-Dimensional Data

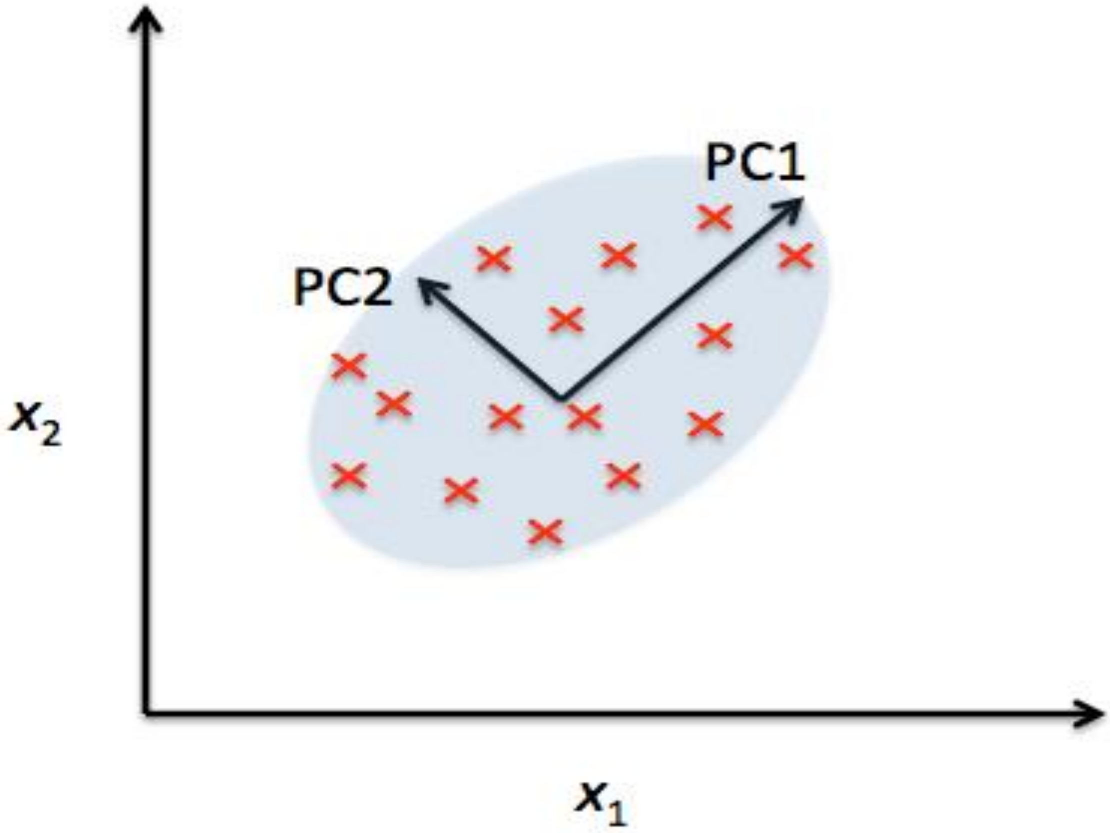
- Training a model with high-dimensional data requires much time-space complexity
- Overfitting
- Not all the features of the data are relevant to the problem being solved
- Data in lower dimension has lower noise(unnecessary parts of the data)

Type of Dimensionality Reduction

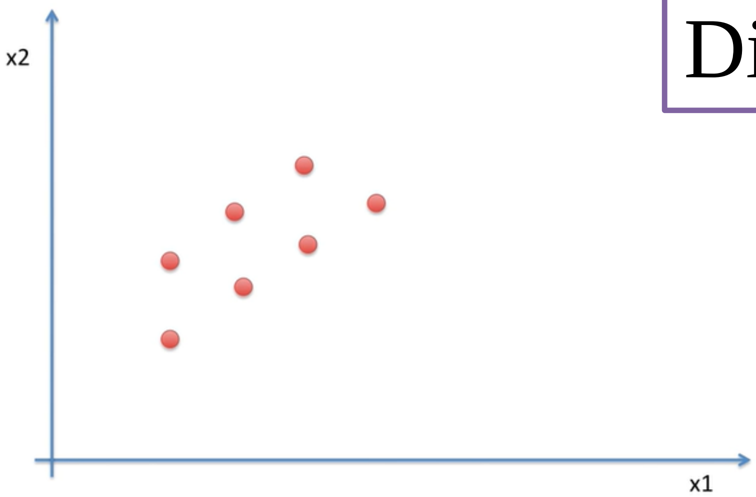
Feature Extraction: This technique has to do with finding new features in the data after it has been transformed from a high-dimensional space to a low dimensional space.

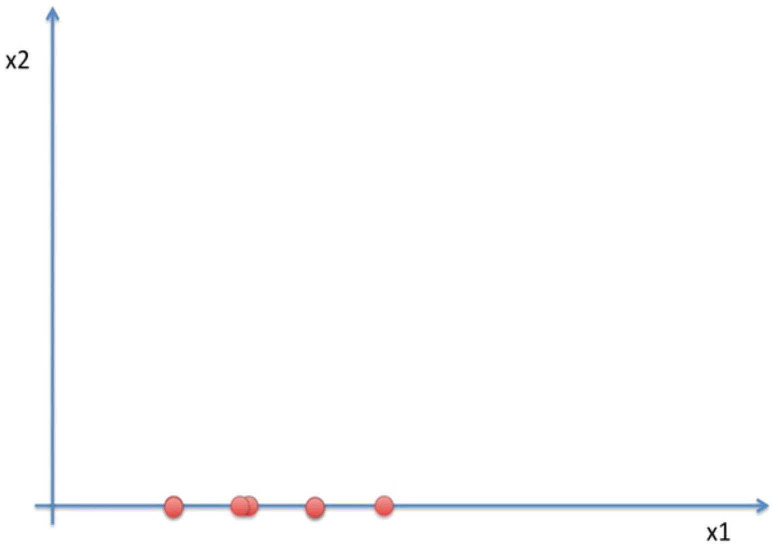
Feature Selection: This have to do with finding the most relevant features to a problem. This is done by obtaining a subset or key features of the original variables

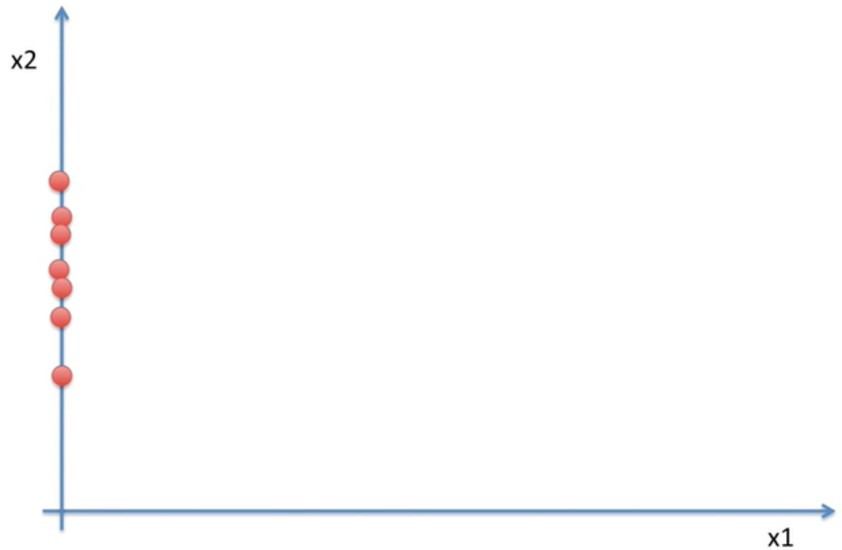
PCA



Dimensionality reduction







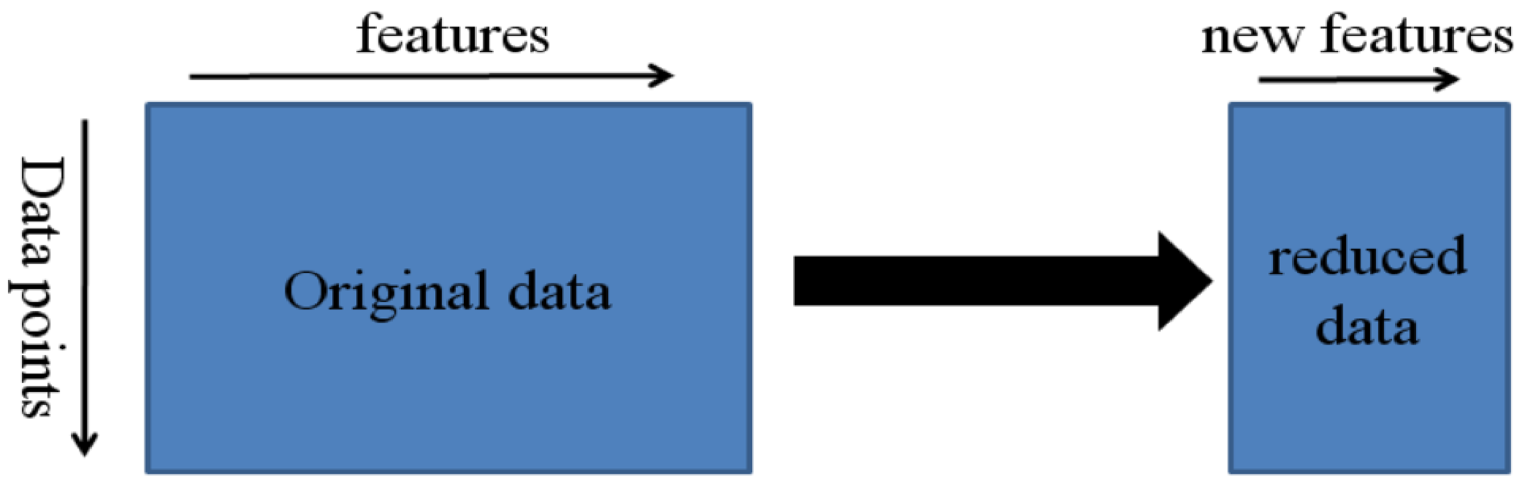
Principal Component Analysis

As the dimension of data increases ,the difficulty to visualize it and perform computation on it also increases.

Principal Component Analysis is statistical approach that aims at dimensionality reduction.

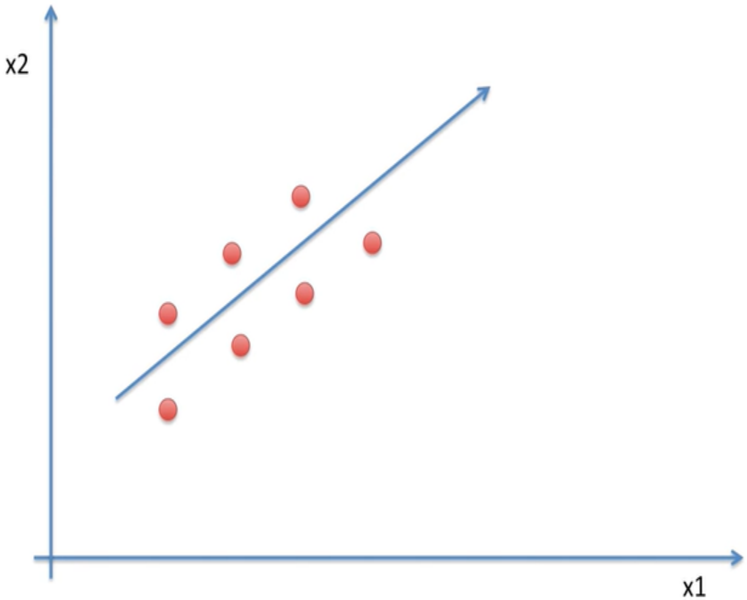
Principal Component Analysis helps us to **identify patterns in data based on the correlation between features**.

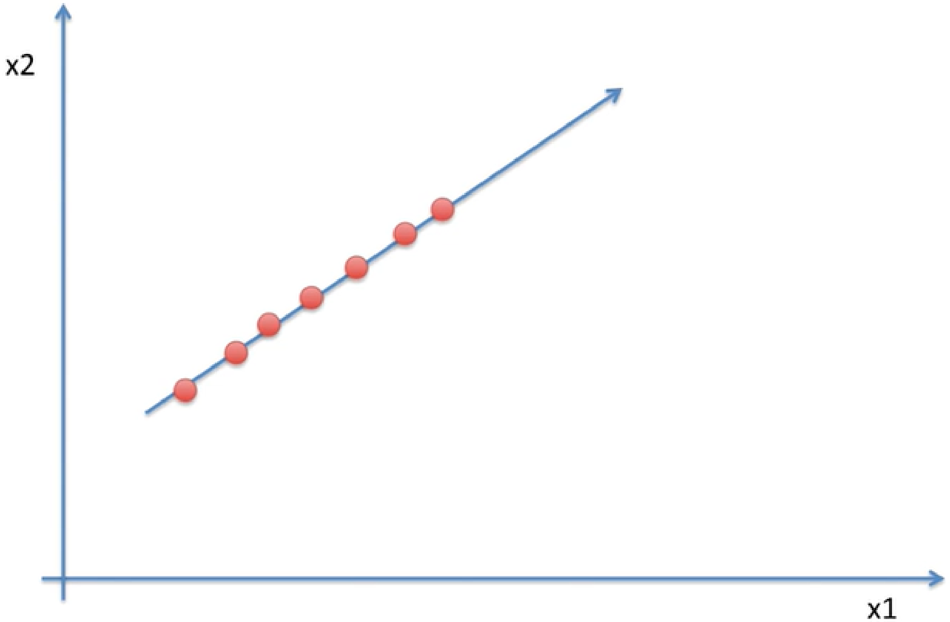
PCA is a method for reducing the number of dimensions in the vectors in a dataset. Essentially, you're compressing the data by exploiting correlations between some of the dimensions.

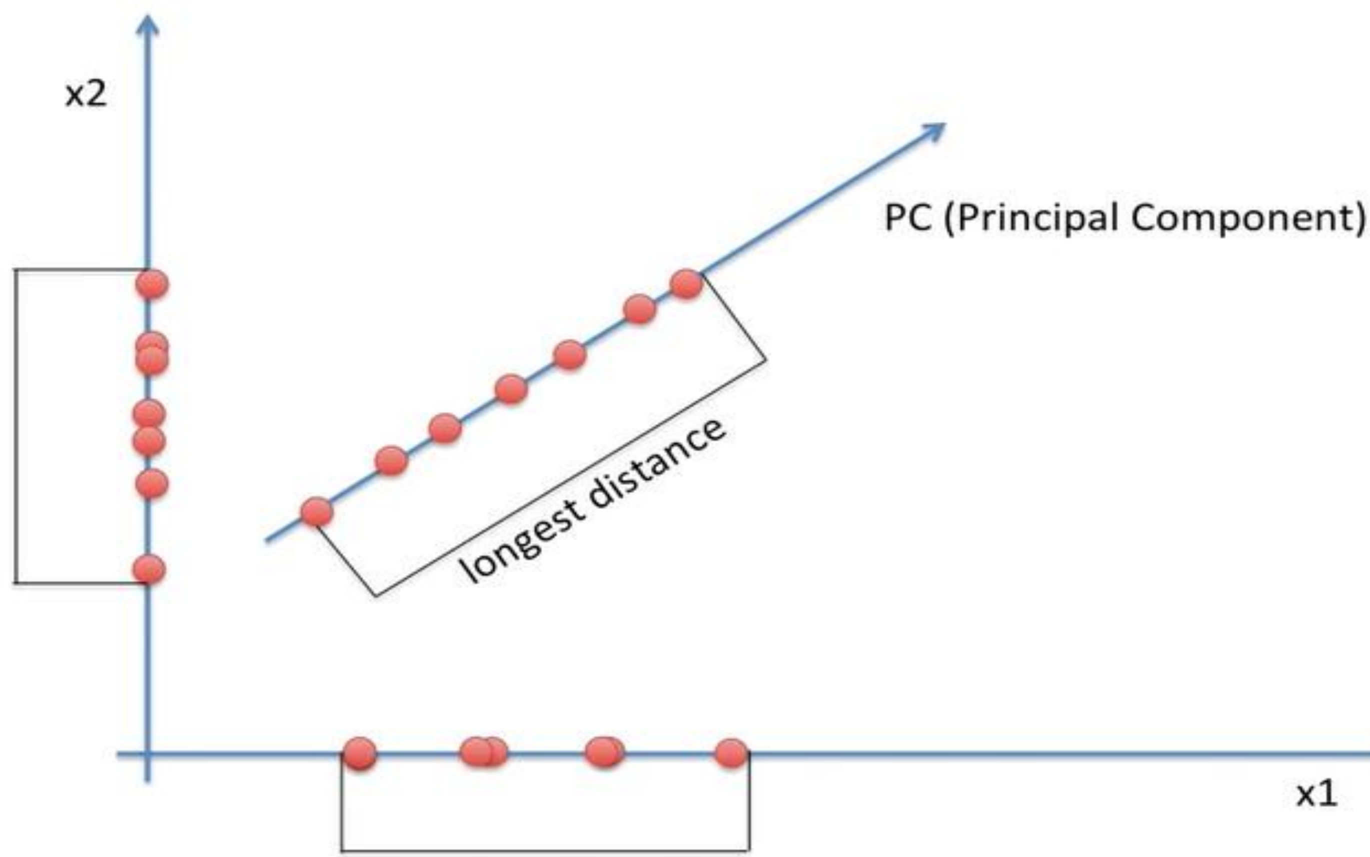


PCA is used to reduce a large number of correlated variables into a smaller set of uncorrelated variables called principal components.

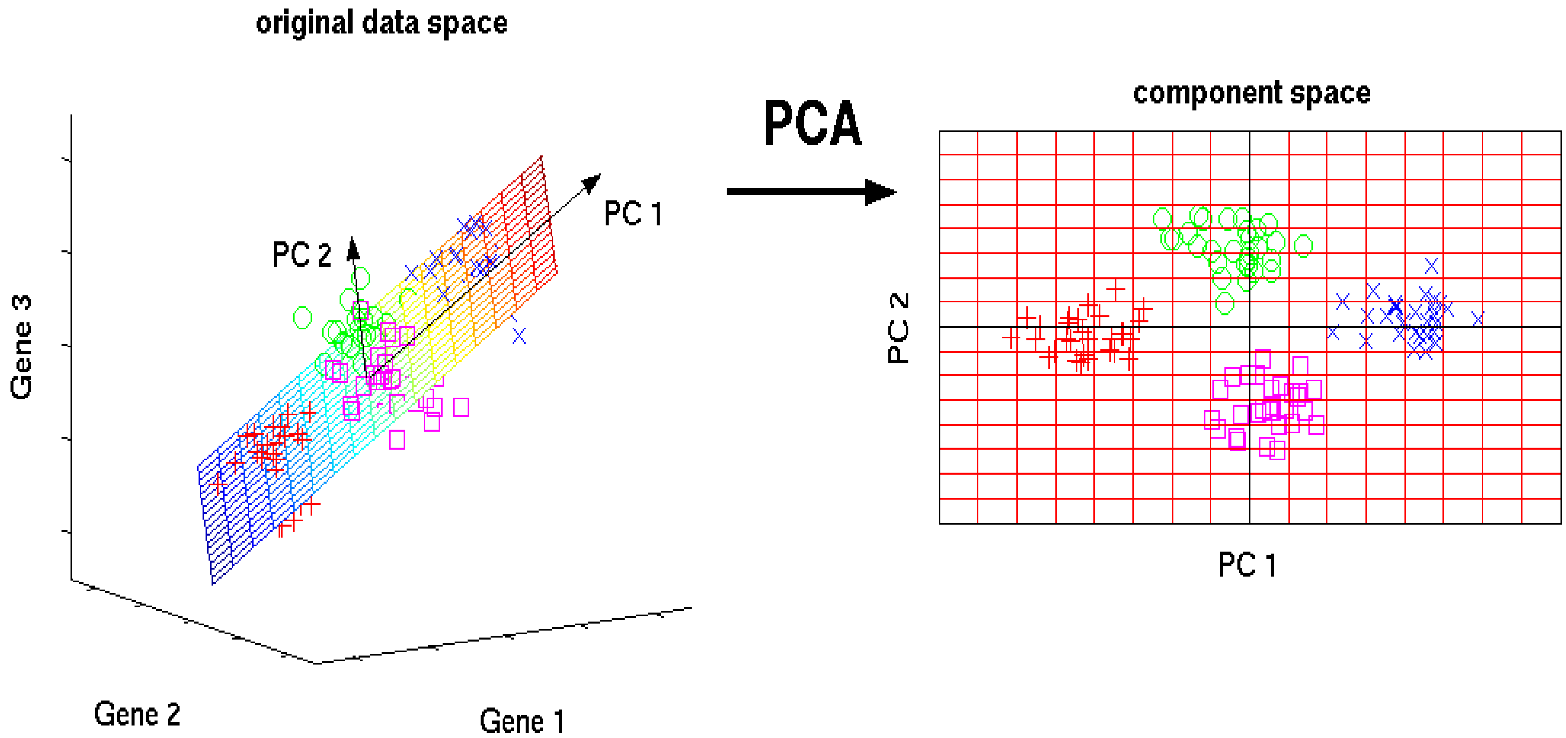
Reduce The Dimension







PC: Eigen Vectors from covariance matrix

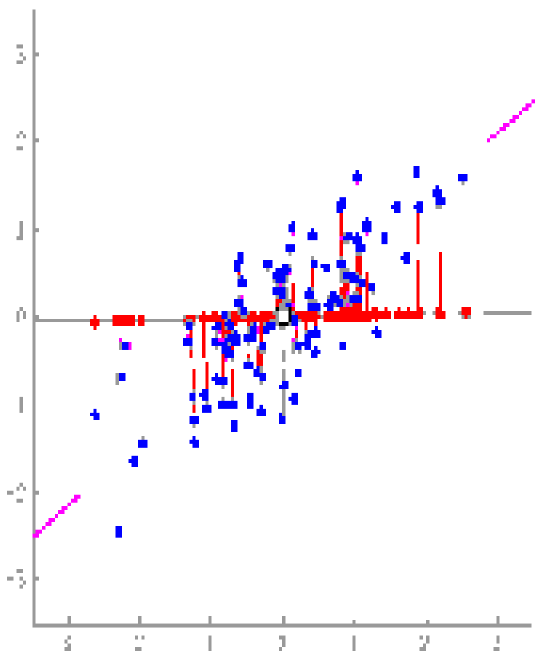


Feature extraction is a process of transforming the data in the high-dimensional space to a space of fewer dimensions.

PCA uses "**orthogonal linear transformation**" to project the features of a data set onto a new coordinate system where the feature which explains the most variance is positioned at the first coordinate (thus becoming the first principal component).

PCA allows us to determine which features capture similar information and discard them to create a more parsimonious model.

PCA is a **variance-maximizing** technique that projects the original data onto a direction that maximizes variance. PCA performs a linear mapping of the original data to a lower-dimensional space such that the variance of the data in the low-dimensional representation is maximized.



preliminaries

- Matrices
- Vectors
- Standard deviations
- Variance
- Co-variance
- Matrix and vector multiplications
- Orthogonality
- Eigenvalues and eigenvectors

Mean and Standard deviation

Mean tells about the middle data element whereas the deviation talks about the spread of data.

Mean= $$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Standard deviation = $$\sqrt{\frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n-1}}$$

Variance

Variance gives the spread of data. The only difference is in the square-root of variance gives us the deviation.

$$\sigma^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$$

Variance is a measure of the variability or spread in a set of data. Mathematically, it is the average squared deviation from the mean score.

Co-variance

Variance measures the deviation from the mean points in one dimension.

Co-variance : It measures how much of the dimensions varies from mean with respect to each other.

How interrelated two data set are.

$$\operatorname{cov}(X,Y) = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{n-1}$$

Co-variance values

- **+ve value:** Both of them increase or decrease together.
- **-ve value:** If one increases, the other decreases.
- **Zero value:** The dimensions are independent. Absolutely no relation.

$$C = \begin{pmatrix} Cov(X,X) & Cov(X,Y) & Cov(X,Z) \\ Cov(Y,X) & Cov(Y,Y) & Cov(Y,Z) \\ Cov(Z,X) & Cov(Z,Y) & Cov(Z,Z) \end{pmatrix}$$

co-variance matrix

Co-variance matrix contain variance of dimension as the main diagonal.
Thus the matrix is symmetric about the main diagonal

$$\mathrm{cov}(X,Y)=\mathrm{cov}(Y,X)$$

$$Matrix(Covariance) = \begin{bmatrix} Var[X_1] & Cov[X_1, X_2] \\ Cov[X_2, X_1] & Var[X_2] \end{bmatrix}$$

Student	Math	English	Art
1	90	60	90
2	90	90	30
3	60	60	60
4	60	60	90
5	30	30	30

$$\mathbf{A} = \begin{bmatrix} 90 & 60 & 90 \\ 90 & 90 & 30 \\ 60 & 60 & 60 \\ 60 & 60 & 90 \\ 30 & 30 & 30 \end{bmatrix}$$

First, we transform the raw scores in matrix A to deviation scores in matrix a.

Then, to find the deviation score sums of squares matrix, we compute a'a,

$$\boxed{a=A-(IA/N)}$$

$$\mathbf{a} = \begin{bmatrix} 90 & 60 & 90 \\ 90 & 90 & 30 \\ 60 & 60 & 60 \\ 60 & 60 & 90 \\ 30 & 30 & 30 \end{bmatrix} - \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 90 & 60 & 90 \\ 90 & 90 & 30 \\ 60 & 60 & 60 \\ 60 & 60 & 90 \\ 30 & 30 & 30 \end{bmatrix} (1/5)$$

$$\mathbf{a} = \begin{bmatrix} 90 & 60 & 90 \\ 90 & 90 & 30 \\ 60 & 60 & 60 \\ 60 & 60 & 90 \\ 30 & 30 & 30 \end{bmatrix} - \begin{bmatrix} 66 & 60 & 60 \\ 66 & 60 & 60 \\ 66 & 60 & 60 \\ 66 & 60 & 60 \\ 66 & 60 & 60 \end{bmatrix}$$

$$\mathbf{a} = \begin{bmatrix} 24 & 0 & 30 \\ 24 & 30 & -30 \\ -6 & 0 & 0 \\ -6 & 0 & 30 \\ -36 & -30 & -30 \end{bmatrix}$$

To find the deviation score sums of squares matrix, we compute **a'a**

$$\mathbf{a'a} = \begin{bmatrix} 24 & 24 & -6 & -6 & -36 \\ 0 & 30 & 0 & 0 & -30 \\ 30 & -30 & 0 & 30 & -30 \end{bmatrix} \begin{bmatrix} 24 & 0 & 30 \\ 24 & 30 & -30 \\ -6 & 0 & 0 \\ -6 & 0 & 30 \\ -36 & -30 & -30 \end{bmatrix}$$

$$\mathbf{a'a} = \begin{bmatrix} 2520 & 1800 & 900 \\ 1800 & 1800 & 0 \\ 900 & 0 & 3600 \end{bmatrix}$$

To create the variance-covariance matrix, we divide each element in the deviation sum of squares matrix by n

$$\mathbf{V} = \mathbf{a}'\mathbf{a} / n$$

$$\mathbf{V} = \begin{bmatrix} 2520/5 & 1800/5 & 900/5 \\ 1800/5 & 1800/5 & 0/5 \\ 900/5 & 0/5 & 3600/5 \end{bmatrix}$$

$$\mathbf{V} = \begin{bmatrix} 504 & 360 & 180 \\ 360 & 360 & 0 \\ 180 & 0 & 720 \end{bmatrix}$$

Eigenvalues and eigenvectors

Let A be an nxn matrix. The number is an eigenvalue of A if there exists a non-zero vector v such that $Av = \lambda v.$
In this case, vector v is called an eigenvector of A corresponding to .

An eigenvector is a vector whose direction remains unchanged when a linear transformation is applied to it.

$$Matrix\ A * vector\ V = \text{Resultant R}$$

$$\begin{bmatrix} 2 & 3 \\ 2 & 1 \end{bmatrix} \times \begin{bmatrix} 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 12 \\ 8 \end{bmatrix} = 4 \times \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$

$$\mathbf{A} \;.\; \mathbf{v} \qquad = \lambda \;.\; \mathbf{v}$$

When the matrix – called as transformational matrix is multiplied by Eigenvector, result is another vector that is transformed from its original position

Eigen vectors

- For n*n square matrix, it will have n eigen vectors
- Even if the vector is scaled by some amount prior to multiplication, the same multiple is obtained as the result. This is because we are not changing the direction and just making it long.
- **All eigenvectors of the matrix are perpendicular to each other: orthogonal**

eigenvalues

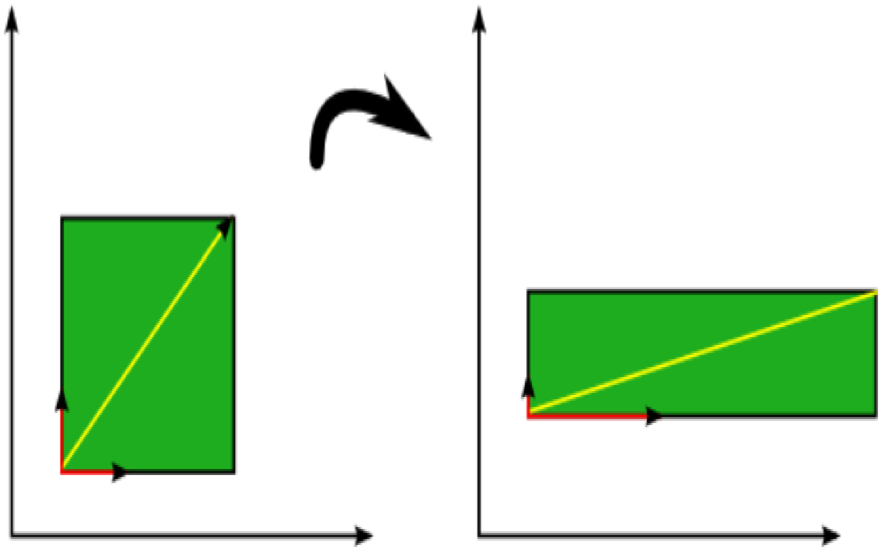
$$A.v = \lambda.v$$
$$\equiv A.v - \lambda.I.v = 0$$
$$\equiv (A - \lambda.I).v = 0$$

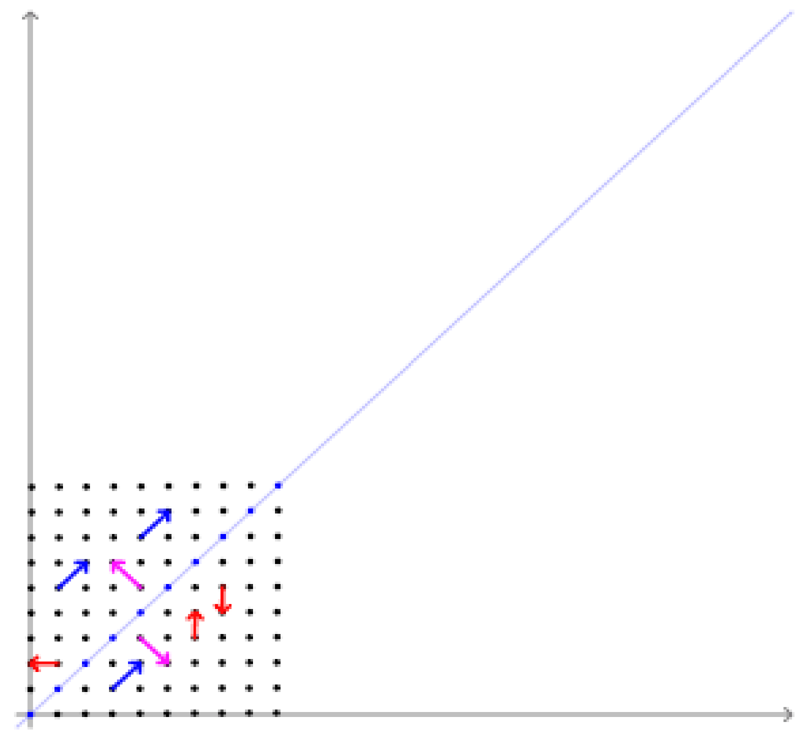
$$A = \begin{bmatrix} 2 & 3 \\ 2 & 1 \end{bmatrix}.$$

$$Det \begin{pmatrix} 2-\lambda & 3 \\ 2 & 1-\lambda \end{pmatrix} = 0.$$

$$\begin{aligned} &(2-\lambda)(1-\lambda) - 6 = 0 \\ \Rightarrow &2 - 2\lambda - \lambda - \lambda^2 - 6 = 0 \\ \Rightarrow &\lambda^2 - 3\lambda - 4 = 0. \end{aligned}$$

$$\begin{aligned} \lambda_1 &= \frac{-b-\sqrt{D}}{2a} = \frac{3-5}{2} = -1, \\ \lambda_2 &= \frac{-b+\sqrt{D}}{2a} = \frac{3+5}{2} = 4. \end{aligned}$$





Principal Component Analysis

- PCA is the vector who maximize the variance
- Minimize the correlation

The main idea of principal component analysis is to reduce the dimensionality of the data set consisting of many variable correlated to each other.

The principal components are the eigenvector of covariance matrix and hence they are orthogonal.

Step for PCA

1.Calculate the covariance matrix x of Data Point

$$\begin{bmatrix} V_a & C_{a,b} & C_{a,c} & C_{a,d} & C_{a,e} \\ C_{a,b} & V_b & C_{b,c} & C_{b,d} & C_{b,e} \\ C_{a,c} & C_{b,c} & V_c & C_{c,d} & C_{c,e} \\ C_{a,d} & C_{b,d} & C_{c,d} & V_d & C_{d,e} \\ C_{a,e} & C_{b,e} & C_{c,e} & C_{d,e} & V_e \end{bmatrix}$$

$$\hat{\Sigma} = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{x}_i - \hat{\boldsymbol{\mu}})(\mathbf{x}_i - \hat{\boldsymbol{\mu}})^\top$$

2.Calculate the eign vector and corresponding eign value.

3.Sort the eign vector accordingly to their eign value In decreasing order

4.Choose first K eign vector and that will be the new K dimensions,

5.Transform the origin n dimension data point into K dimesions.

[CovarienceMatrix].[EigenVector]=[EigenValue][EignVector]

The overall goal of PCA is to reduce the number of d dimensions (features) in a dataset by projecting it onto a k dimensional subspace where $k < d$

1. Standardize the data.
2. Use the standardized data to generate a covariance matrix (or perform Singular Vector Decomposition).
3. Obtain eigenvectors (principal components) and eigenvalues from the covariance matrix. Each eigenvector will have a corresponding eigenvalue.
4. Sort the eigenvalues in descending order.
5. Select the k eigenvectors with the largest eigenvalues, where k is the number of dimensions used in the new feature space ($k \leq d$).
6. Construct a new matrix with the selected k eigenvectors.