Principal Component Analysis

Omendra gangwar (226150101)

Mehta Family School of Data Science and Artificial Intelligence

Indian Institute of Technology Guwahati
Guwahati - 781039, Assam

August 18, 2023

Presentation Overview

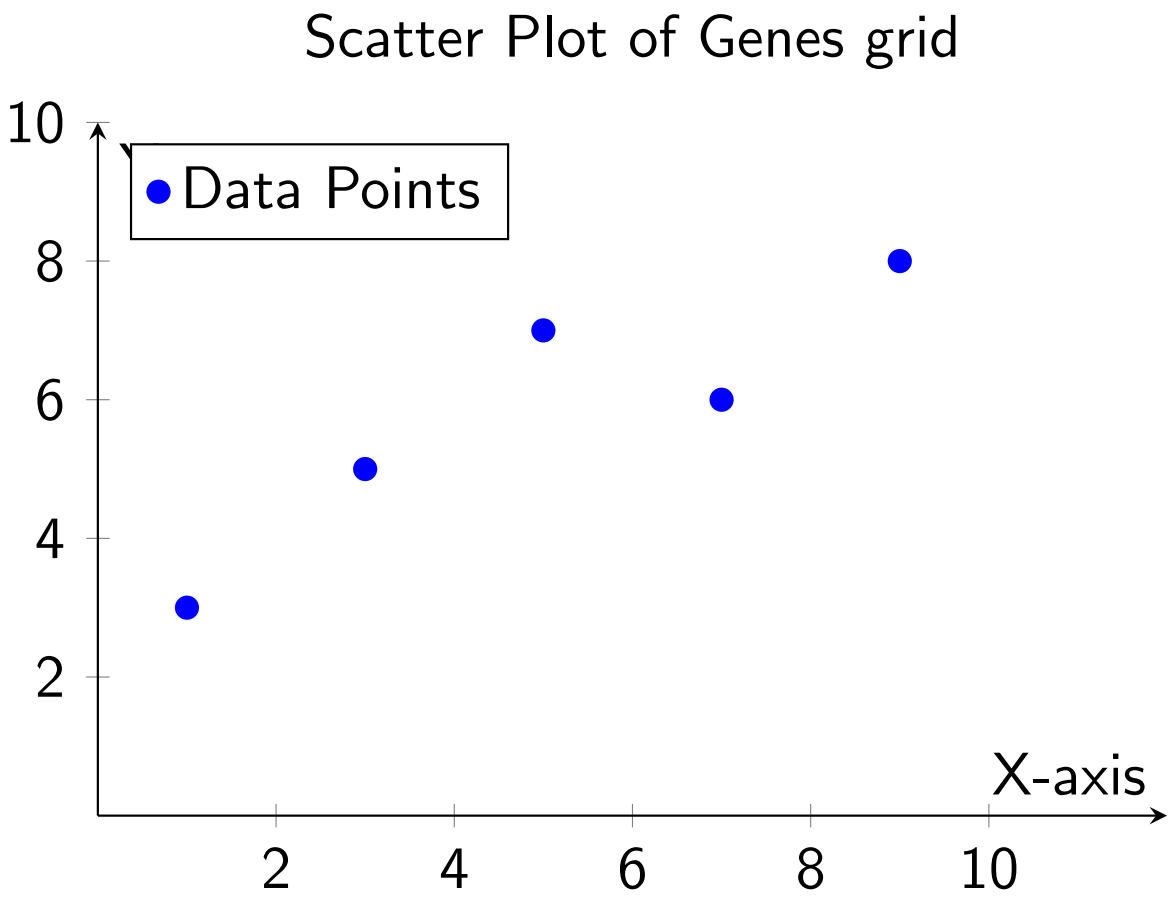
Gene Expression Data

Suppose we have a dataset of gene expression measurements for five different samples and number of genes are two.

	Gene 1	Gene 2
Sample 1	$E_{1,1}$	$E_{2,1}$
Sample 2	$E_{1,2}$	$E_{2,2}$
Sample 3	$E_{1,3}$	$E_{2,3}$
Sample 4	$E_{1,4}$	$E_{2,4}$
Sample 5	$E_{1,5}$	$E_{2,5}$

Table: Gene Expression Data

scatter plot

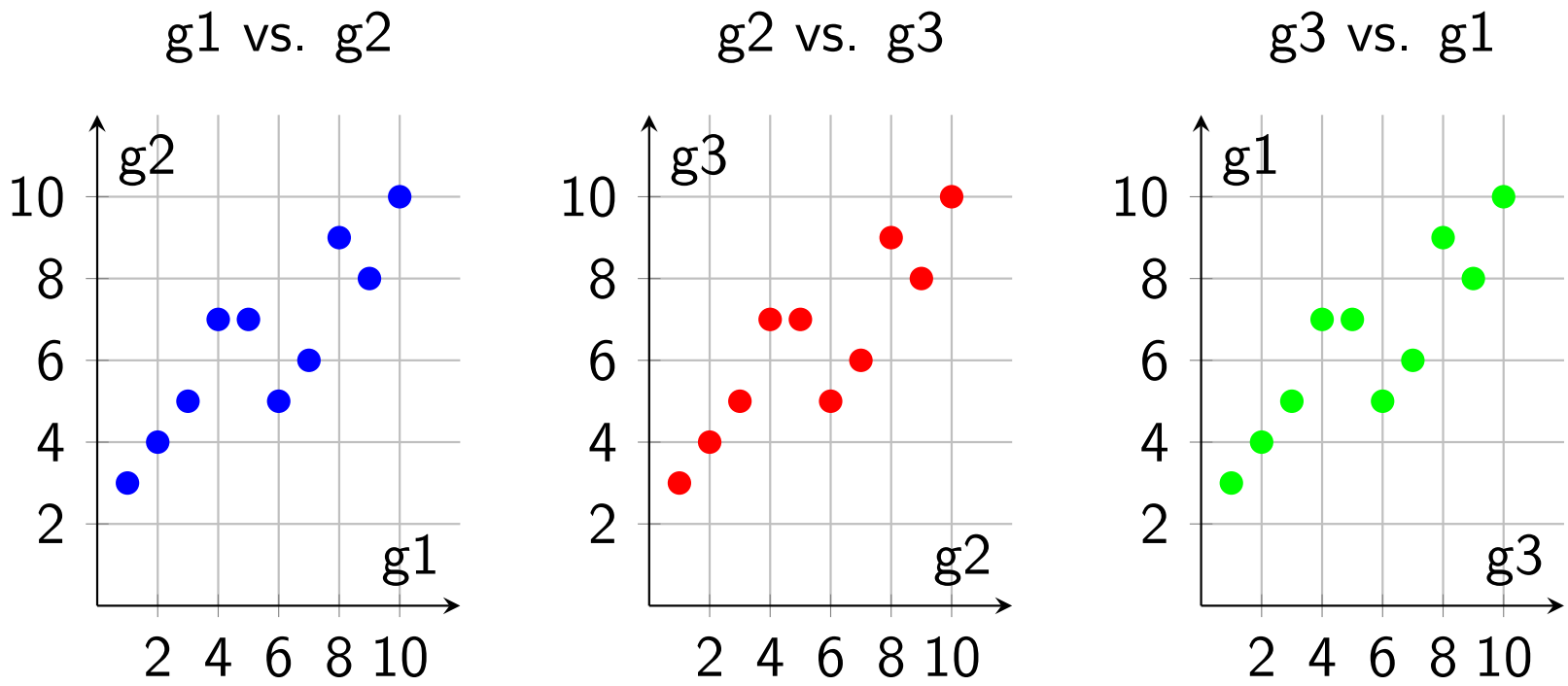


If we incease the observations

	g1	g2	g3
Sample 1	E1,1	E2,1	E3,1
⋮	⋮	⋮	⋮
Sample 10	E1,10	E2,10	E3,10

Table: Gene Expression Data

Scatter Plots



Big and High-Dimensional Data

Image Compression

1. Images are inherently high-dimensional data
2. Each pixel in an image corresponds to a dimension
3. For color images, each pixel might have multiple color channels
4. grayscale image has 256 x 256 = 65,536 dimensions, and a 256x256 RGB image has 256 x 256 x 3 = 196,608 dimensions.

Key Problems in Higher dimension data

1. How to make meaningful Visualization?
2. How to do meaningful quantitative information?

Dimensionality Reduction

In machine learning, there are several techniques for reducing the dimensionality of data.

1. **Principal Component Analysis (PCA)**
2. t-Distributed Stochastic Neighbor Embedding (t-SNE)
3. Autoencoders
4. Linear Discriminant Analysis (LDA)

If you want to explore more click it.

Principal Component Analysis (PCA)

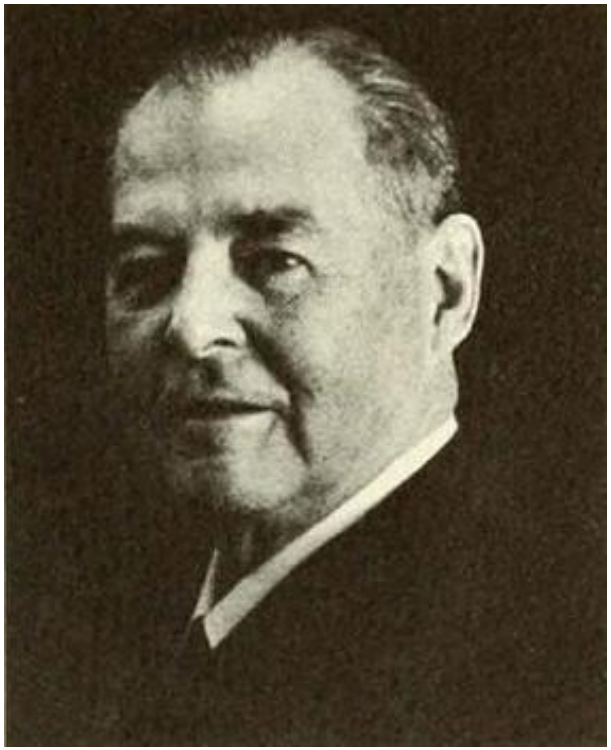
PCA

Early Developments (Late 19^{th} Century)

1. The origins of PCA can be traced back to the late 19th century (1901) with the work of Karl Pearson, a British mathematician and statistician. He introduced the concept of "principal axes" and the idea of reducing multivariate data to lower-dimensional representations.
2. Pearson's work laid the groundwork for understanding variance and covariance in datasets.

Emergence of Principal Component Analysis (PCA) (1930s-1940s)

1. The term "Principal Component Analysis" was coined by Harold Hotelling, an American statistician, in the 1933s.
2. Hotelling's work extended Pearson's concepts, emphasizing the importance of covariance matrices and the idea of finding orthogonal axes (principal components) that capture the most variance in the data..



What is PCA ?

1. Principal component analysis, or PCA, is a technique that is widely used for applications such as dimensionality reduction, lossy data compression, feature extraction. and data visualization
2. Unsupervised technique for extracting variance structure from high dimensional datasets.
3. It is also known as the **Karhunen-Loève transform**

Definition in Bishop book

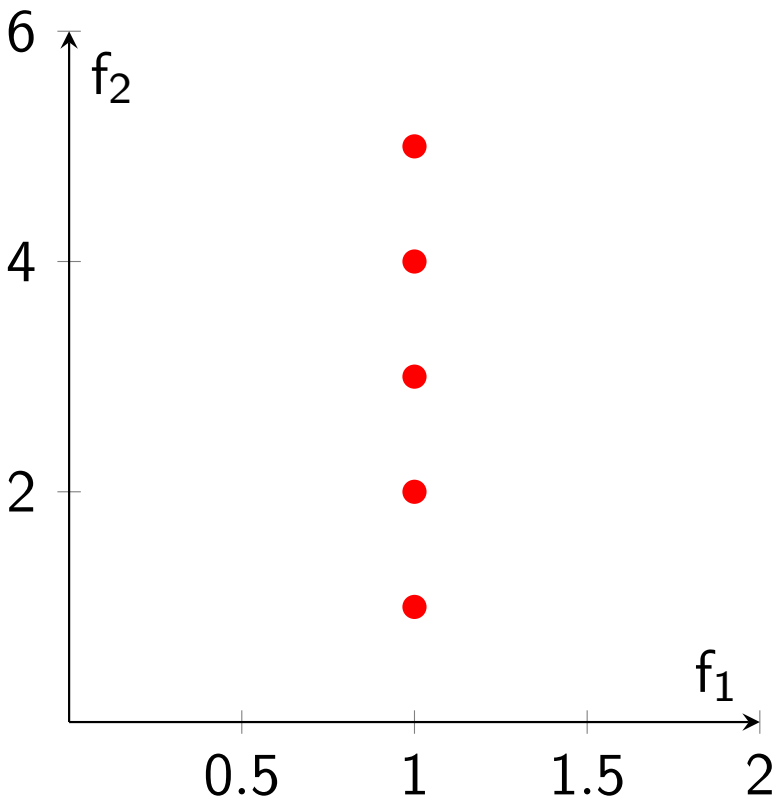
Hotelling, 1933

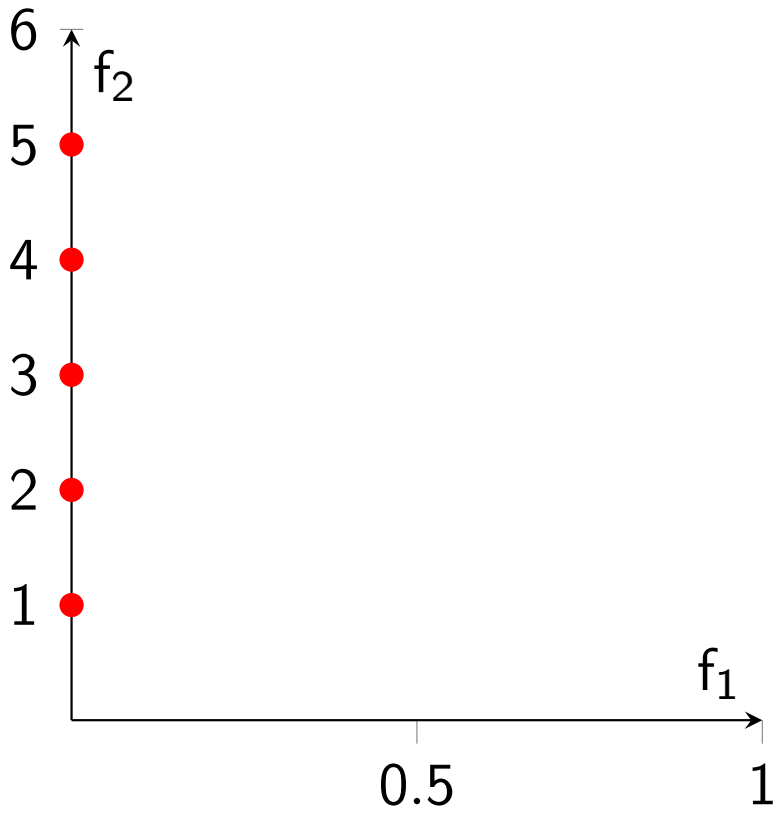
PCA can be defined as the orthogonal projection of the data onto a lower dimensional linear space, known as the principal subspace, such that the variance of the projected data is maximized .

Pearson, 1901

Equivalently, it can be defined as the linear projection that minimizes the average projection cost, defined as the mean squared distance between the data points and their projections.

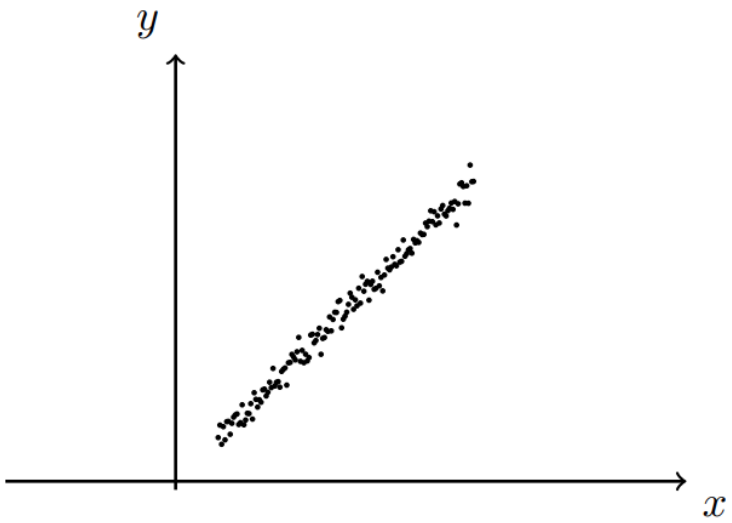
Geometric Intuition of PCA

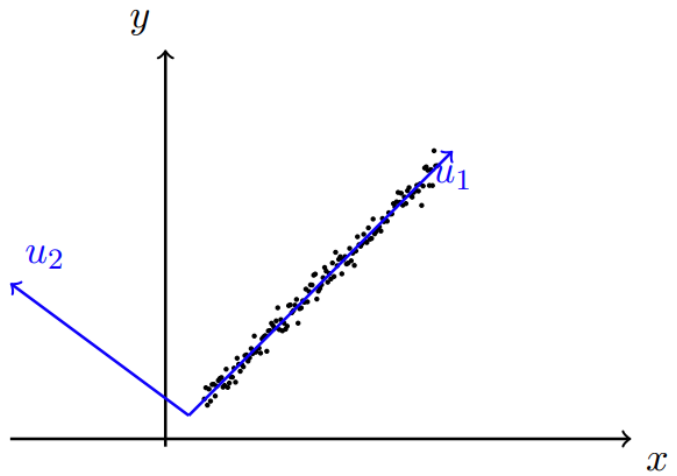




Geometric Intuition of PCA

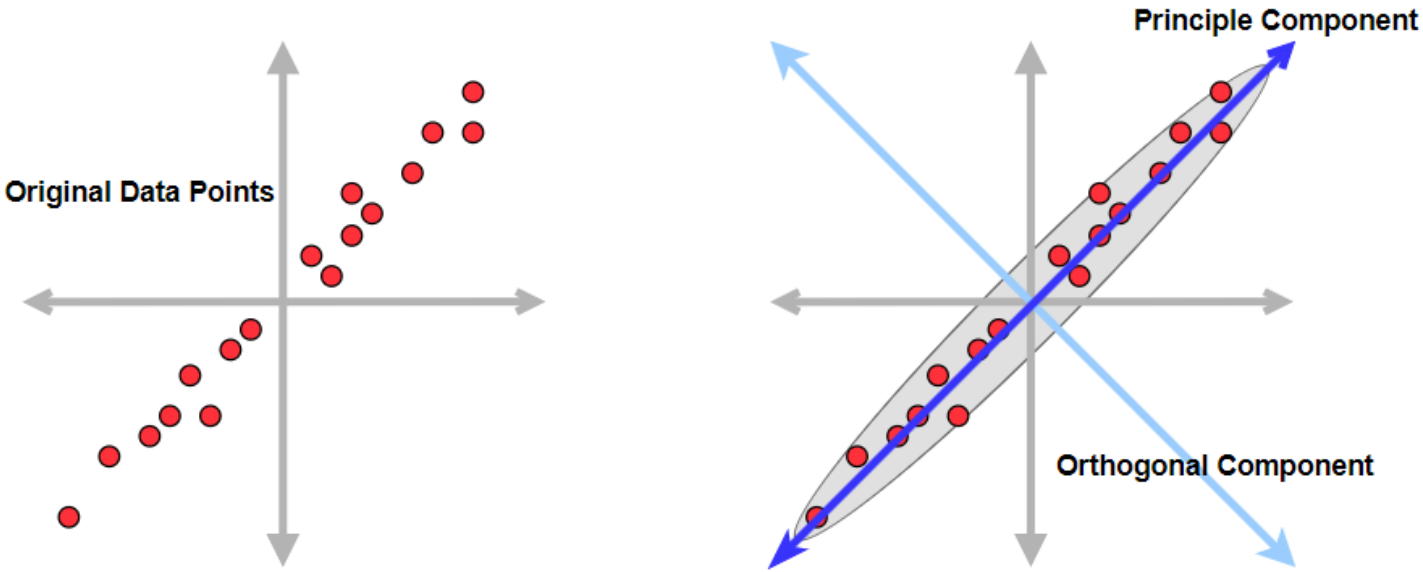
Let's our data is in 2-d we want in lower dimensional space to 1-d.





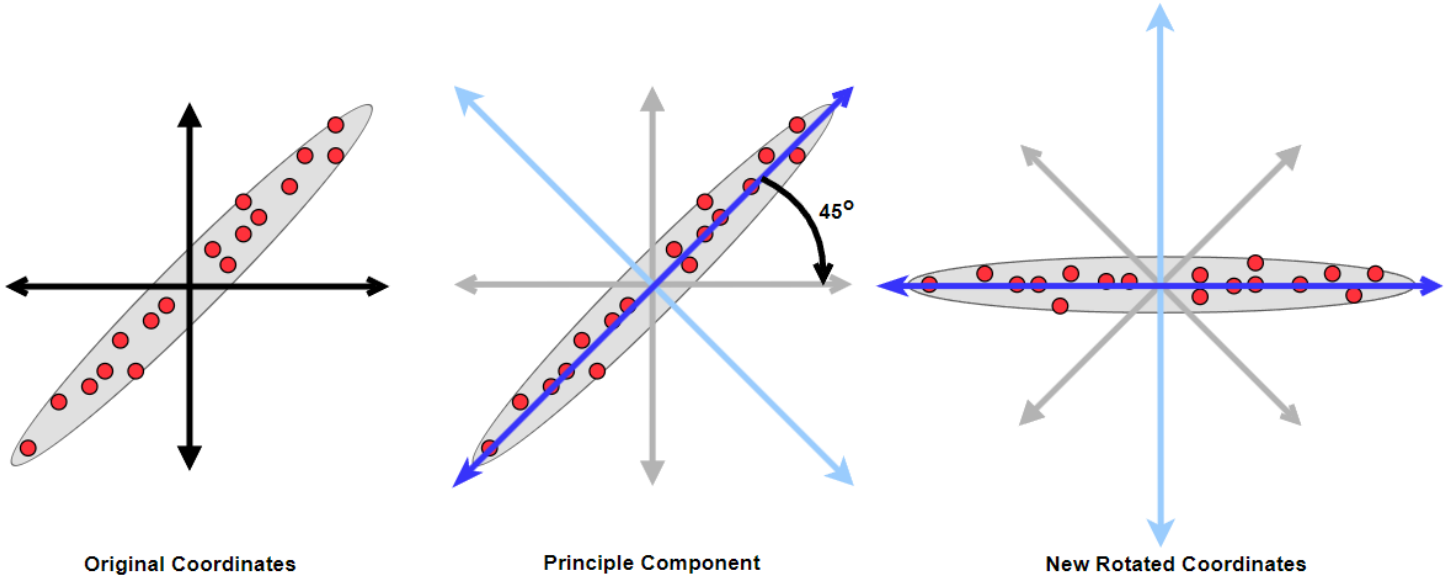
Geometric Intuition of PCA

Assume we have X=2-d dataset and X has been already column standardized with mean zero and standard deviation=1



Here spread in f1 and f2 feature are same. Now which feature would select based on maximum variance. So we rotate the axis such both new axis are Perpendicular

Centerd data



PCA:Two Interpretations

Maximum Variance Direction:

Let $\mathbf{X}$ be the data matrix with dimensions $N \times D$, where N is the number of samples and D is the original feature dimension. The goal is to find the direction $\mathbf{u}$ that maximizes the variance when the data is projected onto this direction.

Mathematical formulation:

$$\mathbf{u}_{\text{max}} = \arg\max_{\mathbf{u}} \left(\frac{1}{N} \sum_{n=1}^{N} (x_n \cdot \mathbf{u})^2 \right)$$

where $\mathbf{u}_{\text{max}}$ is the direction vector that maximizes the variance of the projected data.

Minimum Reconstruction Error:

Minimum Reconstruction Error:

1st PC a vector u such that projection on to this vector yields minimum MSE reconstruction:

$$\sum_{i=1}^{n} \|\mathbf{x}_i - (\mathbf{u}^T \mathbf{x}_i)\mathbf{u}\|^2$$

Maximum Variance Formulation

Consider a data set of observations $\{x_n\}$ where $n = 1, \ldots, N$, and x_n is a Euclidean variable with dimensionality D. Our goal is to project the data onto a space having dimensionality $M < D$ while maximizing the variance of the projected data. For the moment, we shall assume that the value of M is given.

$$\mathbf{X_{NxD}} \rightarrow \mathbf{T_{NxM}}$$

Let the projection onto a one-dimensional space (M = 1). We can define the direction of this space using a D-dimensional vector u_1, and we shall choose to be a unit vector so that $||u|| = 1$. Now,each data point X_n is then projected onto a scalar value $u_1^\top x_n$.

Projection of data

2-dimensional data:

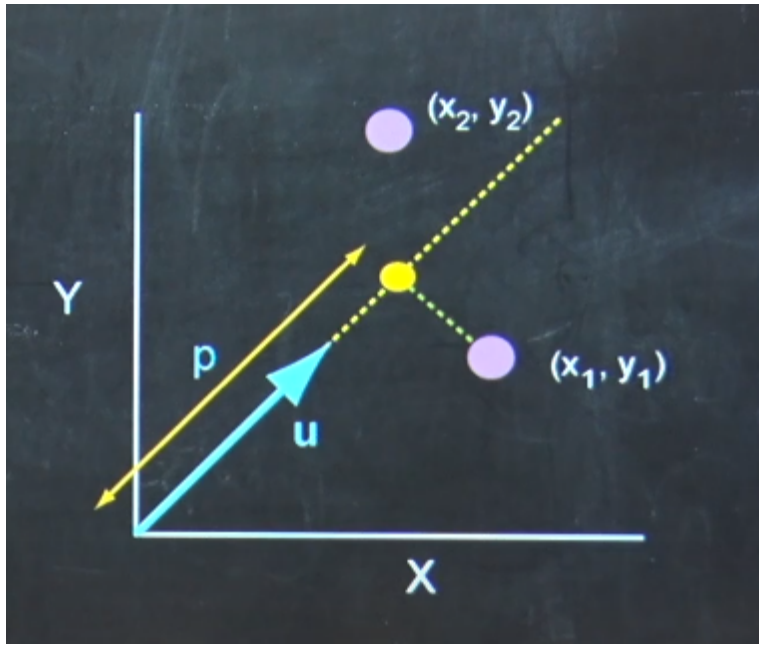
$$X = \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \end{bmatrix}$$

u is a unit vector: $\|\mathbf{u}\| = 1$

Data point 1:

$$X_1 = \begin{bmatrix} X_1 \\ Y_1 \end{bmatrix}$$

$$p = \text{proj}_{\mathbf{u}}\mathbf{X}_1 = \mathbf{X}_1 \cdot \mathbf{u}$$



continue

The mean of the projected data is $u_1^\top \bar{x}$ where $\bar{x}$ is the sample set mean given by

$$\bar{x} = \frac{1}{N}\sum_{n=1}^{N} x_n$$

Our Task: Find u_1 such that the variance of the projected data is maximum.

The variance of the projected data is given by-

$$var(u_1^\top x_n) = \frac{1}{N}\sum_{n=1}^{N}\{u_1^\top x_n - u_1^\top \bar{x}\}^2 = u_1^\top S u_1$$

where S is the data covariance matrix defined by

$$\mathbf{S} = \frac{1}{N}\sum_{n=1}^{N}(\mathbf{x}_n - \bar{\mathbf{x}})(\mathbf{x}_n - \bar{\mathbf{x}})^{\mathrm{T}}$$

PCA: an optimization problem

Data

$$\mathrm{X} = \begin{bmatrix} x_{11} & \cdots & x_{1j} & \cdots & x_{1d} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{i1} & \cdots & x_{ij} & \cdots & x_{id} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nj} & \cdots & x_{nd} \end{bmatrix}$$

Unit vector

$$\mathbf{u} = \begin{bmatrix} v_1 \\ \vdots \\ v_d \end{bmatrix}$$

Projection of X **on** u : $\mathrm{proj}_{\mathbf{u}}\,\mathrm{X} = \mathrm{Xu}$

Objective function: $\mathrm{argmax}_{\mathbf{u}}[\mathrm{var}(\mathbf{Xu})] = \mathrm{argmax}[u_1^\top S u_1]$

Constraint: $\|u\| = 1$

Solve optimization problem

To solve optimization problem we introduce a Lagrange multiplier that we shall denote by λ_1, and then make an unconstrained maximization of

$$L(u_1, \lambda_1) = \mathbf{u}_1^{\mathrm{T}} \mathbf{S} \mathbf{u}_1 + \lambda_1 \left(1 - \mathbf{u}_1^{\mathrm{T}} \mathbf{u}_1\right).$$

By setting the derivative with respect to $\mathbf{u}_1$ equal to zero

$$\frac{\partial L}{\partial u_1} = 2\mathbf{S}\mathbf{u}_1 - 2\lambda_1 \mathbf{u}_1$$

we see that this quantity will have a stationary point when

$$\mathbf{S}\mathbf{u}_1 = \lambda_1 \mathbf{u}_1$$

which says that $\mathbf{u}_1$ must be an eigenvector of $\mathbf{S}$.

continue

If we left-multiply by $\mathbf{u}_1^{\mathrm{T}}$ and make use of $\mathbf{u}_1^{\mathrm{T}}\mathbf{u}_1 = 1$, we see that the variance is given by

$$\mathbf{u}_1^{\mathrm{T}}\mathbf{S}\mathbf{u}_1 = \lambda_1$$

and so the variance will be a maximum when we set $\mathbf{u}_1$ equal to the eigenvector having the largest eigenvalue λ_1. This eigenvector is known as the first principal component

Selection of Principal Components

The principal components

$$\mathrm{S}\mathbf{u} = \lambda\mathbf{u}$$

S is a $\mathrm{d} \times \mathrm{d}$ matrix

$$\lambda_i \quad i = 1, 2, \cdots, d \rightarrow \textit{Eigenvalue}$$
$$\mathbf{u}_i \quad i = 1, 2, \cdots, d \rightarrow \textit{Eigenvector}$$

Variance of projected data:

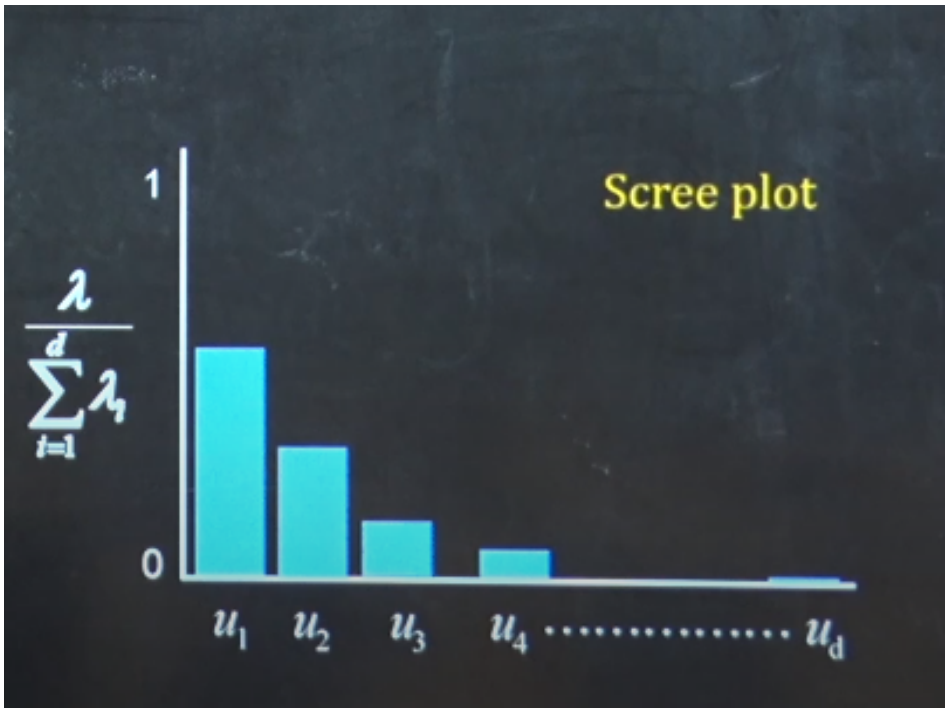
$$\mathrm{var}(\mathbf{Xu}) = \mathbf{u}^T\mathbf{S}\mathbf{u} = \mathbf{u}^T\lambda\mathbf{u} = \lambda\mathbf{u}^T\mathbf{u} = \lambda$$
$$\lambda_1 > \lambda_2 > \cdots > \lambda_d$$

Eigenvector of λ_2 : $\mathbf{u}_2$
Data projected on $\mathbf{u}_2$ will have second highest variance.
$\mathbf{u}_2$ is second principal component.

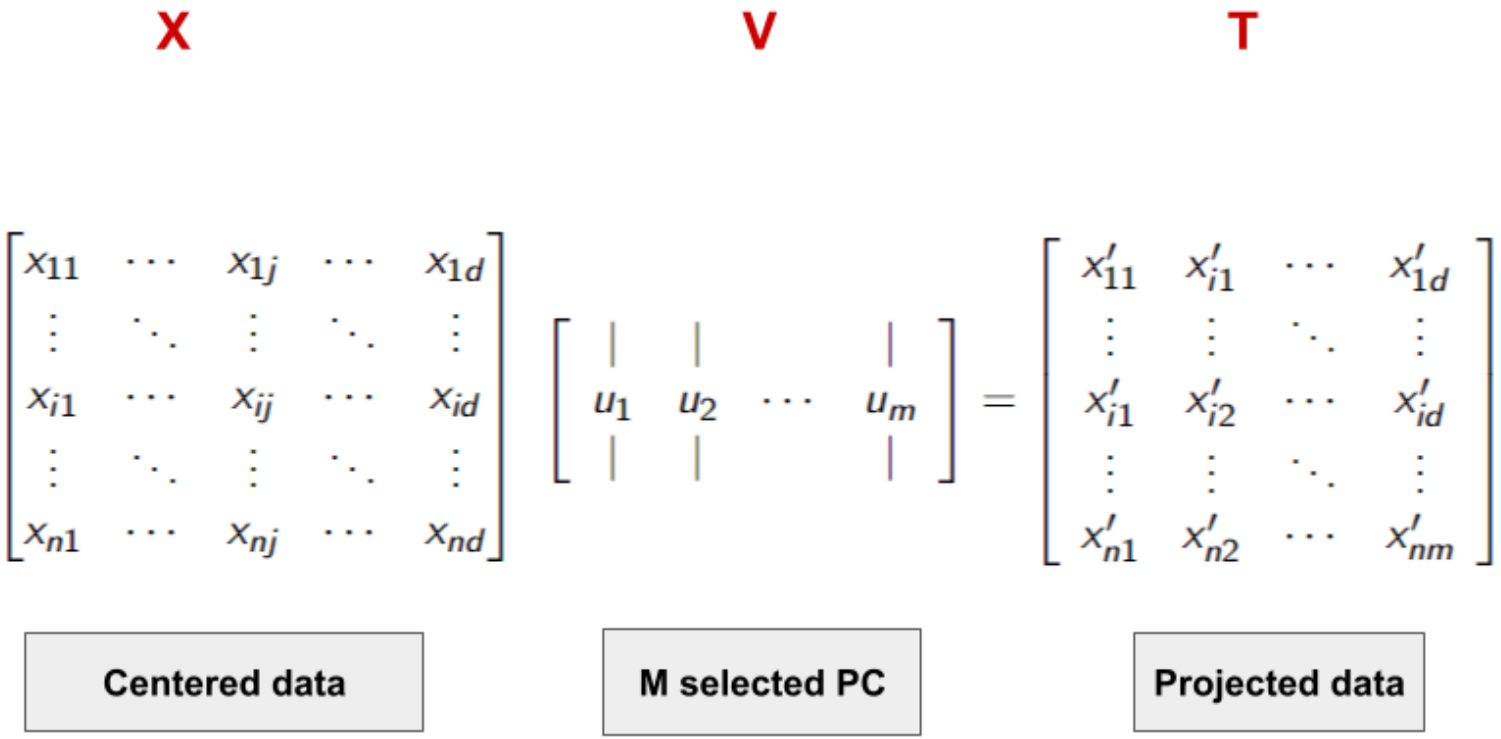
Selection of Principal Components

$$\lambda_1 > \lambda_2 > \cdots > \lambda_d$$

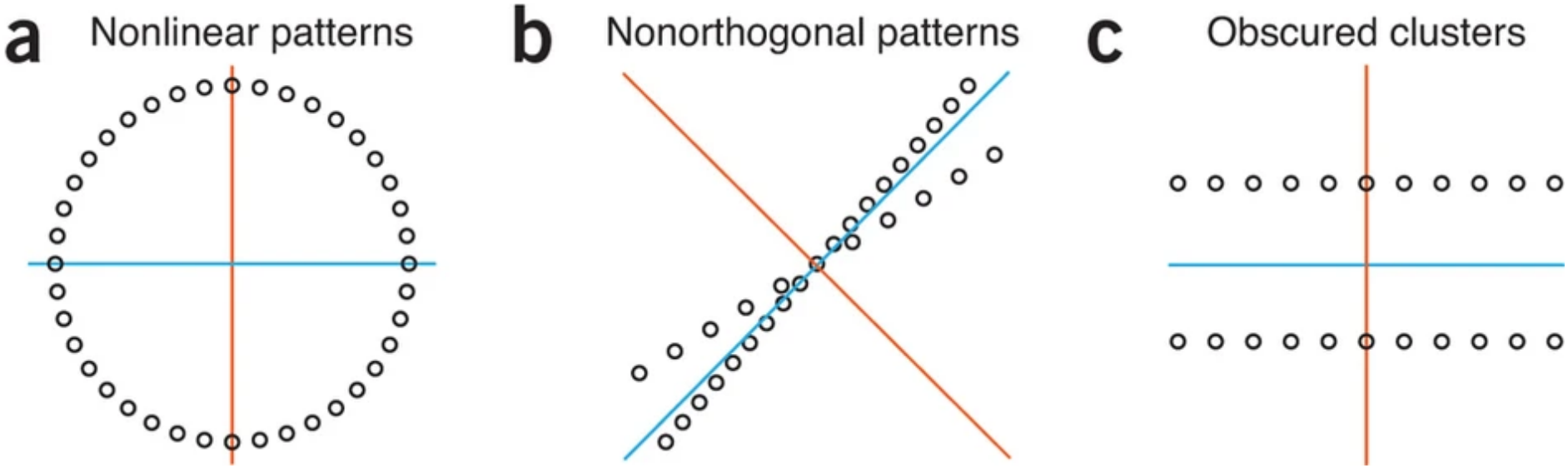
$$\frac{\lambda_1}{\sum_{i=1}^{d} \lambda_i} > \frac{\lambda_2}{\sum_{i=1}^{d} \lambda_i} > \cdots > \frac{\lambda_d}{\sum_{i=1}^{d} \lambda_i}$$



Project data on principal components



Limitation of PCA



References

1. Bishop Chapter-12
2. CS7015: Deep Learning(Lecture-6)
3. Data Analysis for Biologists(Lecture-45)
4. Mathematics for Machine Learning(Chapter-10)
5. Limitation of P.C.A