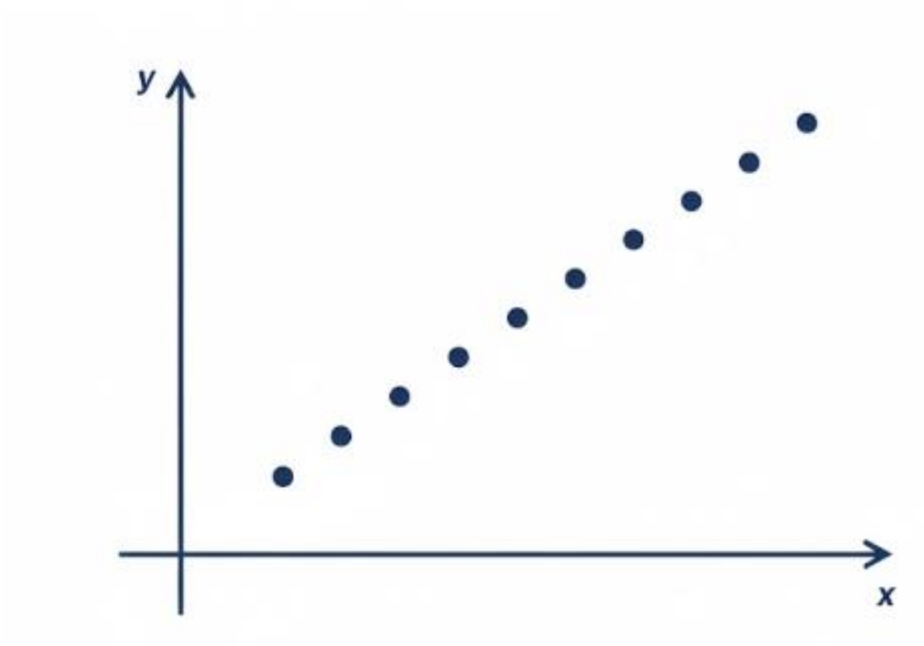
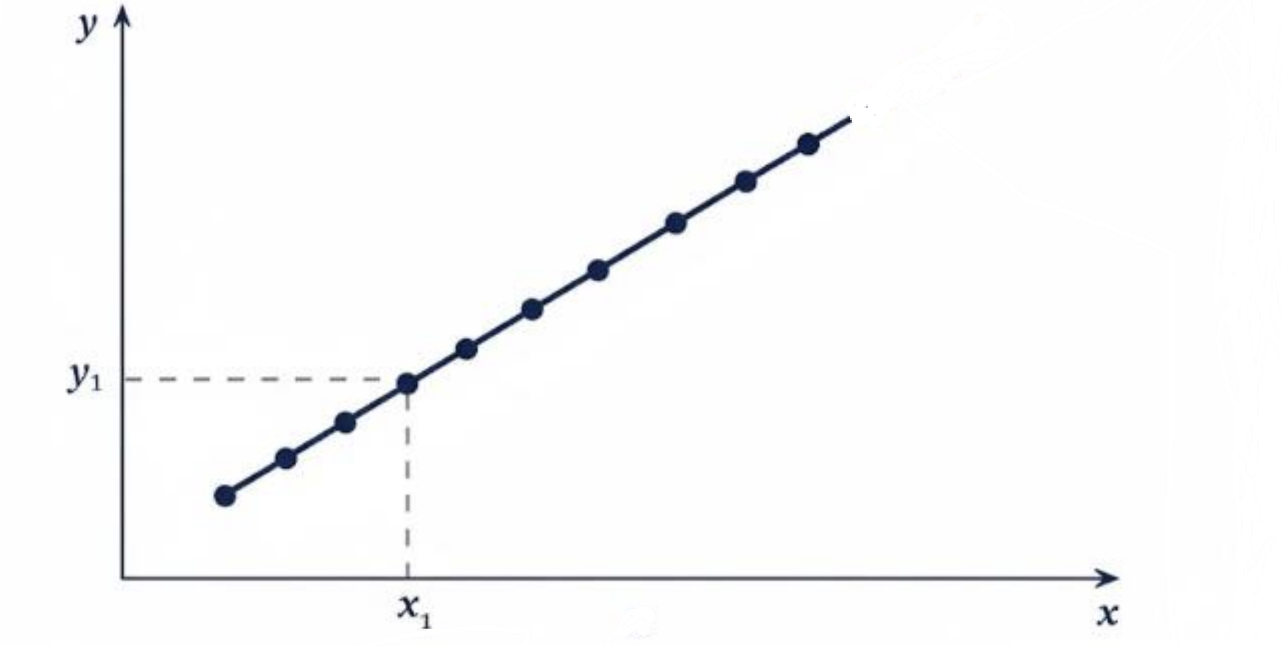
Principal Component Analysis

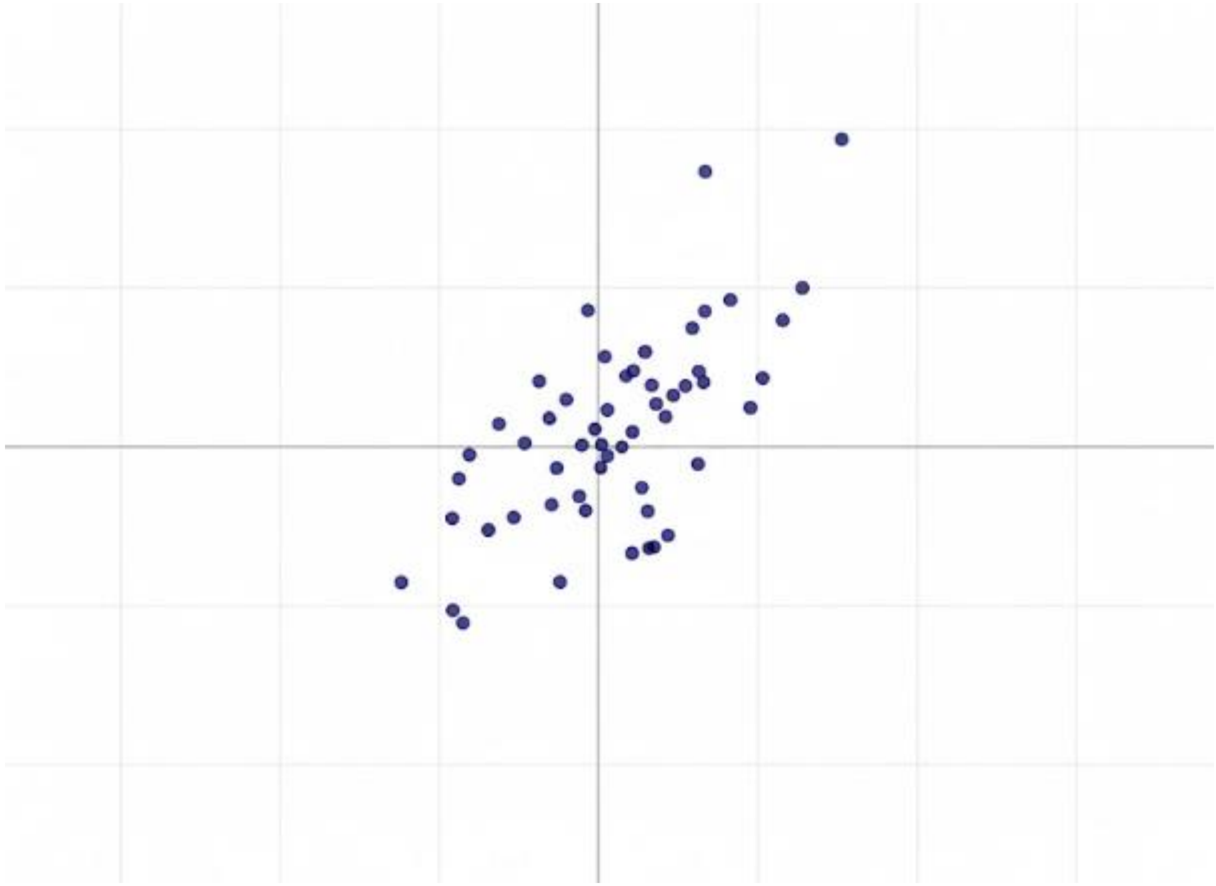
Muhammad Danish Khan

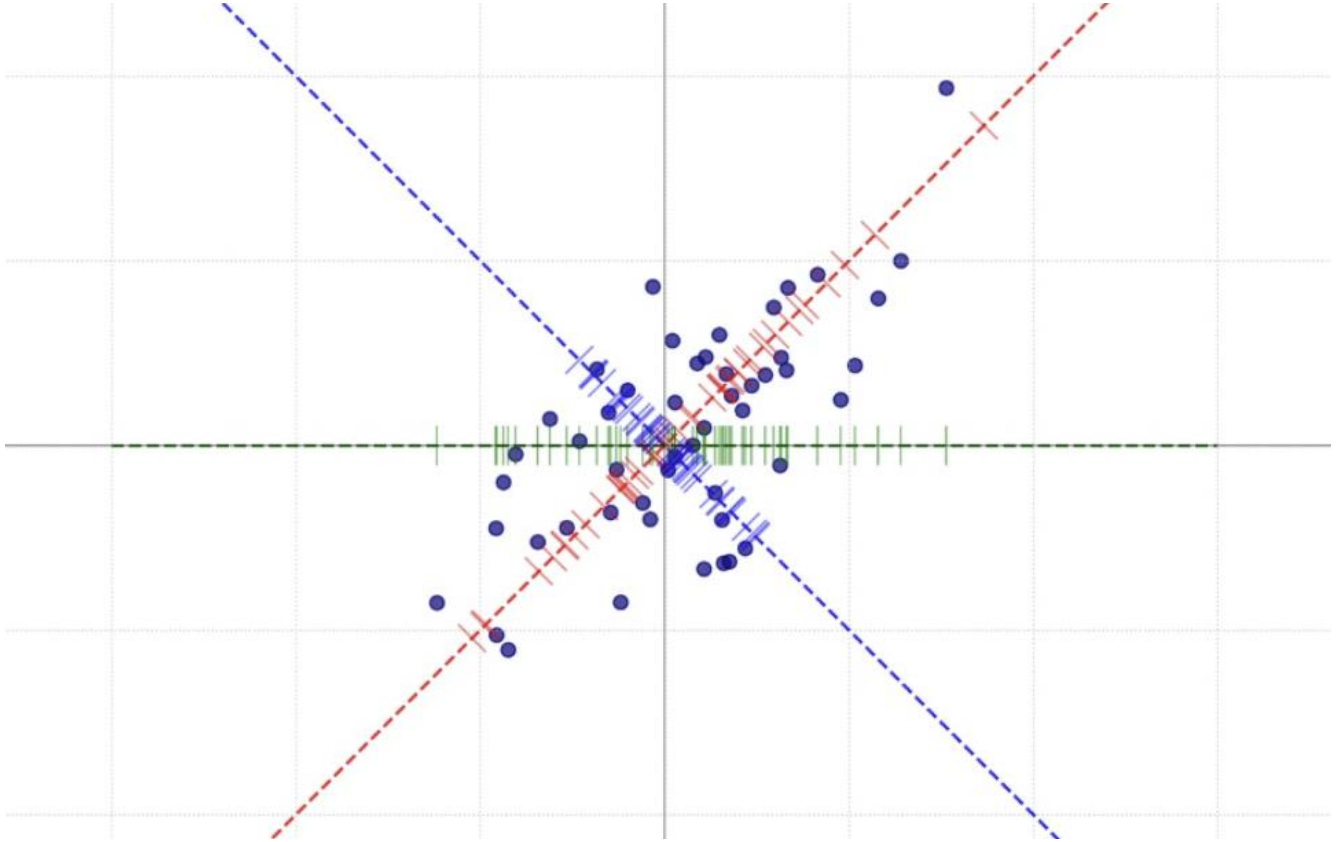
Intuition



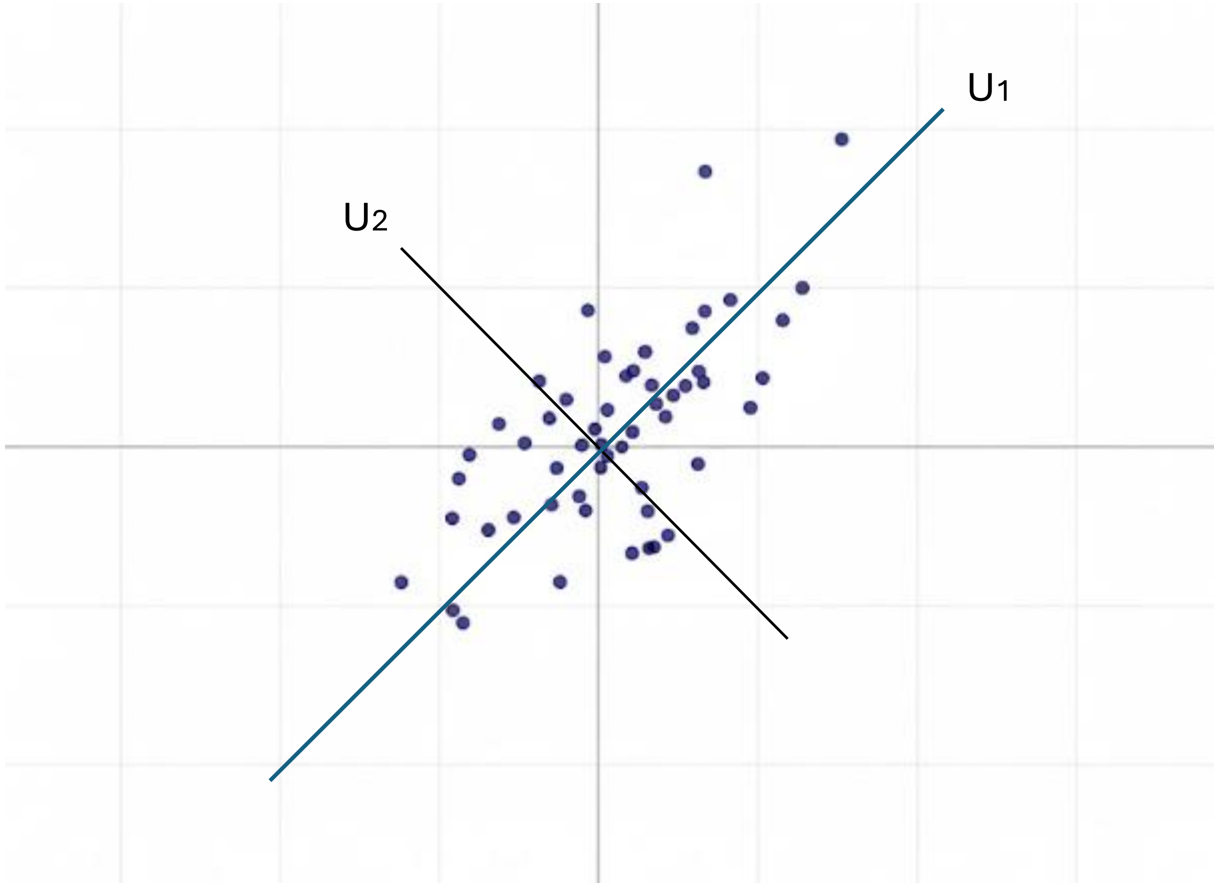


Intuition:





Intuition:



Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a method of dimensionality reduction / feature extraction that transforms the data from a d-dimensional space into a new coordinate system of dimension p, where

$$p \leq d$$

(the worst case would be to have $p = d$).

- **The goal is to preserve as much of the variance in the original data as possible in the new coordinate systems.**
- **Given data on *d* variables, the hope is that the data points will lie mainly in a linear subspace of dimension lower than *d*.**
- **In practice, the data will usually not lie precisely in some lower dimensional subspace.**
- **The new variables that form a new coordinate system are called *principal components* (PCs).**

- PCs are denoted by $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_d$.
- The principal components form a basis for the data.
- Since PCs are orthogonal linear transformations of the original variables, there is at most d PCs.
- Normally, not all of the d PCs are used but rather a subset of p PCs,

$$\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p.$$

- In order to approximate the space spanned by the original data points

$$\mathbf{x} = [x_1, x_2, \ldots, x_d],$$

  we can choose p based on what percentage of the variance of the original data we would like to maintain.

- The first PC, $\mathbf{u}_1$, is called the first principal component and has the maximum variance; thus, it accounts for the most significant variance in the data.
- The second PC, $\mathbf{u}_2$, is called the second principal component and has the second highest variance, and so on, until $\mathbf{u}_d$, which has the minimum variance.

The most common definition of PCA, due to Hotelling, is that for a given set of data vectors

$$\mathbf{x}_i, \quad i = 1, \ldots, n,$$

the p principal axes are those orthonormal axes onto which the variance retained under projection is maximal.

Problem Setup

$$x \in \mathbb{R}^d$$

$$y \in \mathbb{R}^p \quad (p \ll d)$$

$$X = [x_1, x_2, \ldots, x_n] \in \mathbb{R}^{d \times n}$$

$$x_i \in \mathbb{R}^d$$

PCA Objective

Find a direction $u_1 \in \mathbb{R}^d$ that maximizes the variance of the projected data:

$$\max_{u_1} \ \mathrm{Var}(u_1^\top X)$$

Variance Formulation

$$\mathrm{Var}(u_1^\top X) = u_1^\top S u_1$$

$$S \in \mathbb{R}^{d \times d}$$

Sample covariance matrix of X

Optimization Problem

$$\max_{u_1} \ u_1^\top S u_1$$

$$\text{subject to} \quad u_1^\top u_1 = 1$$

$$u_1 \in \mathbb{R}^d$$

Lagrangian

$$\mathcal{L}(u_1, \lambda) = u_1^\top S u_1 - \lambda \left(u_1^\top u_1 - 1 \right)$$

$$\mathcal{L}(u_1, \lambda) = u_1^\top S u_1 - \lambda \left(u_1^\top u_1 - 1\right)$$

Optimality Condition

$$\frac{\partial \mathcal{L}}{\partial u_1} = 2Su_1 - 2\lambda u_1 = 0$$

$$Su_1 = \lambda u_1$$

Eigenvalue Interpretation

$$u_1^\top S u_1 = \lambda(u_1^\top u_1) = \lambda$$

Variance of projection = eigenvalue

Multiple Principal Components

$$u_1, u_2, \ldots, u_p$$

are chosen as the eigenvectors of S corresponding to the p largest eigenvalues

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$$

$$u_i^\top u_j = 0 \quad (i \neq j)$$

Final PCA Projection

$$U_p = [u_1, u_2, \ldots, u_p]$$

$$Y = U_p^\top X$$

Projection onto the subspace spanned by the top p eigenvectors

Reconstruction of Original Data

Given the PCA projection

$$Y = U_p^\top X$$

the reconstruction of the original data from the reduced representation is

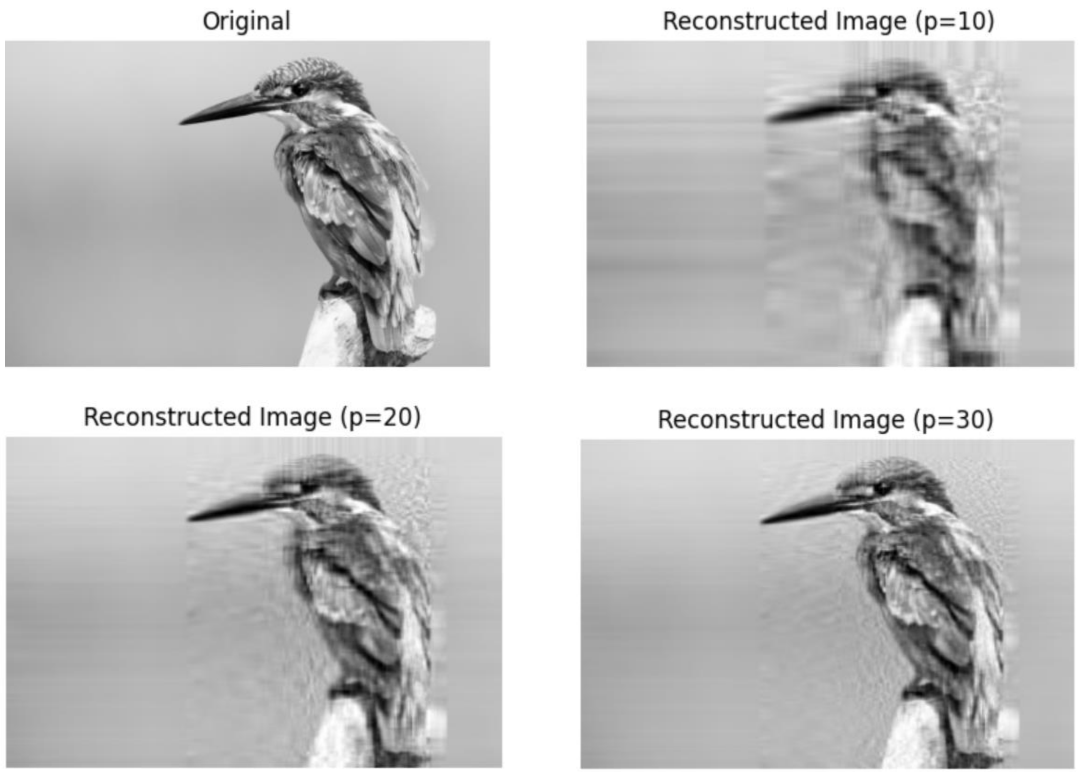
$$\hat{X} = U_p Y$$

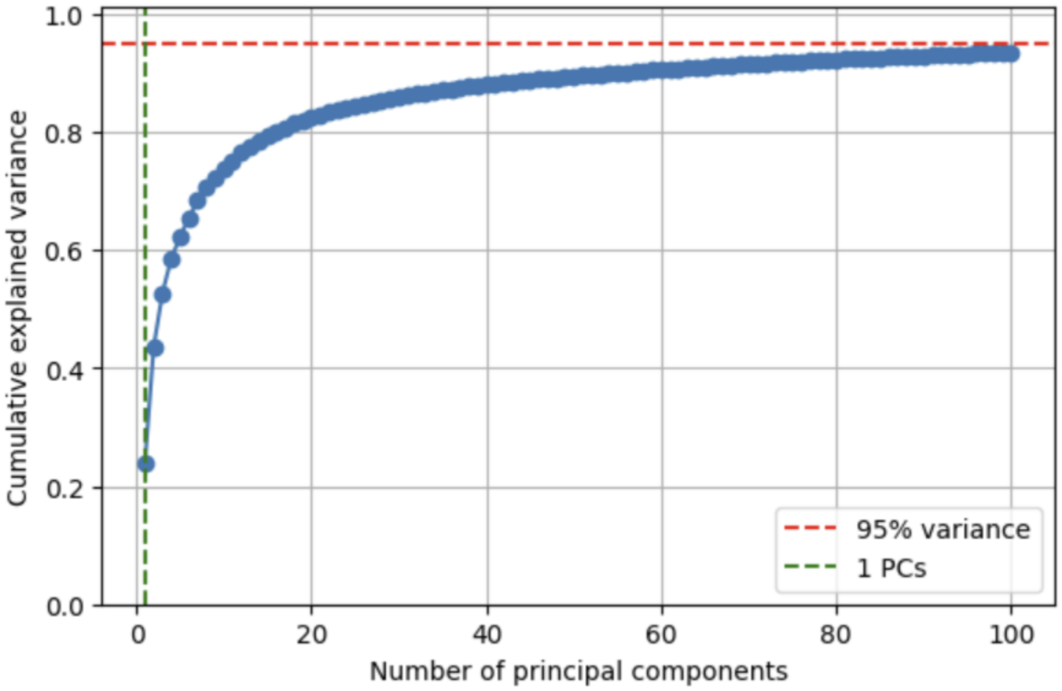
$$\hat{X} = U_p U_p^\top X$$

$$\hat{X} \in \mathbb{R}^{d \times n}$$

This is the best rank-p approximation of X in the least-squares sense.

Application





Thank You