Multivariate Normal and Principal Component Analysis

Gioele La Manno

École Polytechnique Fédérale de Lausanne (EPFL)

School of Life Science (SV)

April 2025

EPFL - BMI - UPLAMANNO

Contents

1 The Multivariate Normal Distribution

2 Principal Component Analysis (PCA)

3 Multiple-Choice Questions

The Multivariate Normal Distribution

From Bivariate to Multivariate Analysis

In our journey through statistical methods, we've progressively increased complexity:

From univariate distributions describing single variables like gene expression levels...
To bivariate distributions examining relationships between pairs of variables...
To regression exploring how multiple predictors inform a single outcome.

Now we take a bold step into multivariate analysis, where we consider many variables simultaneously without designating any specific variable as an outcome.

This marks our entry into unsupervised learning methods - approaches that help us discover patterns and structure without predefined target variables.

The Multivariate Normal Distribution: Definition

The multivariate normal extends the normal distribution to multiple dimensions:

Definition (Multivariate Normal Distribution)

A random vector $\mathbf{X} = (X_1, X_2, \ldots, X_p)^T$ follows a p-dimensional multivariate normal distribution, denoted $\mathbf{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, if its probability density function is:

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right)$$

Where:

- $\boldsymbol{\mu}$ is the mean vector
- $\boldsymbol{\Sigma}$ is the covariance matrix

From Bivariate to Multivariate Normal: Matrix Form

The bivariate normal distribution with correlation ρ is typically written as:

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2} - \frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}\right]\right)$$

We can rearrange the exponent into matrix form:

$$-\frac{1}{2}\begin{pmatrix} x_1-\mu_1 & x_2-\mu_2 \end{pmatrix} \begin{pmatrix} \frac{1}{(1-\rho^2)\sigma_1^2} & \frac{-\rho}{(1-\rho^2)\sigma_1\sigma_2} \\ \frac{-\rho}{(1-\rho^2)\sigma_1\sigma_2} & \frac{1}{(1-\rho^2)\sigma_2^2} \end{pmatrix} \begin{pmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{pmatrix}$$

From Bivariate to Multivariate Normal: Matrix Form (2/2)

This matches the form $-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})$, where:

$$\boldsymbol{\Sigma}^{-1} = \frac{1}{1-\rho^2}\begin{pmatrix} \frac{1}{\sigma_1^2} & \frac{-\rho}{\sigma_1\sigma_2} \\ \frac{-\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{pmatrix}$$

We can find $\boldsymbol{\Sigma}$ by computing the inverse of $\boldsymbol{\Sigma}^{-1}$. First, let's calculate the determinant of $\boldsymbol{\Sigma}^{-1}$:

$$\det(\boldsymbol{\Sigma}^{-1}) = \frac{1}{(1-\rho^2)^2}\left(\frac{1}{\sigma_1^2}\cdot\frac{1}{\sigma_2^2} - \frac{\rho^2}{\sigma_1^2\sigma_2^2}\right) = \frac{1}{(1-\rho^2)\sigma_1^2\sigma_2^2}$$

Now we can calculate $\boldsymbol{\Sigma}$ using the formula for the inverse of a 2 × 2 matrix:

$$\boldsymbol{\Sigma} = (1-\rho^2)\sigma_1^2\sigma_2^2\begin{pmatrix} \frac{1}{\sigma_2^2} & \frac{\rho}{\sigma_1\sigma_2} \\ \frac{\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_1^2} \end{pmatrix}$$

From Bivariate to Multivariate Normal: Matrix Form (3/3)

Simplifying, we get:

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

This reveals that $\boldsymbol{\Sigma}$ is indeed the covariance matrix where:

- Diagonal elements are variances σ_1^2 and σ_2^2
- Off-diagonal elements are covariances, expressed as $\rho\sigma_1\sigma_2$

The determinant is $|\boldsymbol{\Sigma}| = \sigma_1^2\sigma_2^2(1-\rho^2)$, which completes the transformation to the standard multivariate normal form:

$$f(\mathbf{x}) = \frac{1}{2\pi|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

The Covariance Matrix: Encoding Relationships

The covariance matrix $\boldsymbol{\Sigma}$ captures both the spread of individual variables and their interrelationships:

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_p^2 \end{pmatrix}$$

- Diagonal elements $\sigma_i^2 = \text{Var}(X_i)$ represent variable variances
- Off-diagonal elements $\sigma_{ij} = \text{Cov}(X_i, X_j)$ represent covariances

Important Special Cases of Covariance Matrices

Identity Covariance Matrix

When $\mathbf{\Sigma} = \mathbf{I}$, we have the standard multivariate normal with unit variances and uncorrelated variables:

$$\mathbf{\Sigma} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

Important Special Cases of Covariance Matrices

The resulting density has spherical contours in p-dimensional space.

Diagonal Covariance Matrix When off-diagonal elements are zero, variables are uncorrelated but may have different variances.

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_p^2 \end{pmatrix}$$

The resulting density has ellipsoidal contours, with axes aligned with the coordinate axes.

Block Diagonal Covariance Structure

Block Diagonal Covariance Matrix A block diagonal structure indicates groups of correlated variables that are uncorrelated with variables in other groups:

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{A} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{B} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{C} \end{pmatrix}$$

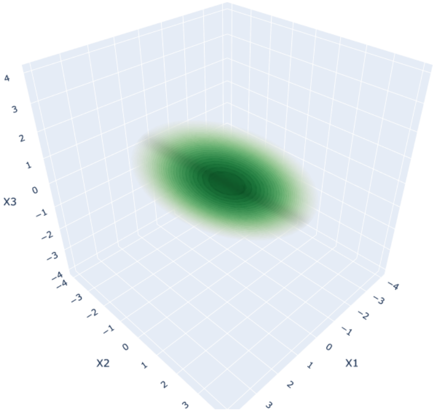
Example (Gene Expression Modules)

In single-cell RNA sequencing, we might observe:

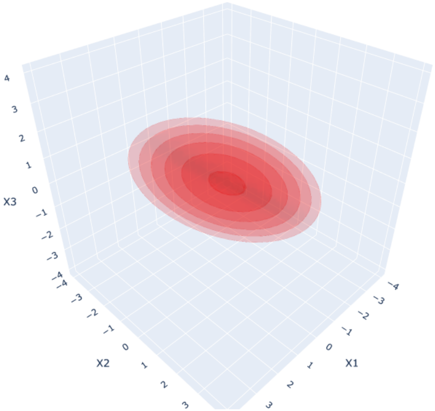
$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{\text{cell cycle}} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{\text{metabolism}} \end{pmatrix}$$

This indicates that cell cycle genes covary with each other but not with metabolism genes.

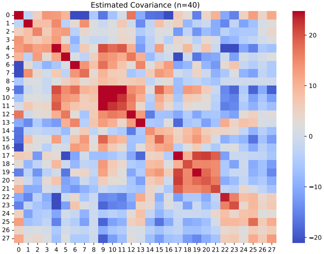
Visualizing the Multivariate Normal Distribution



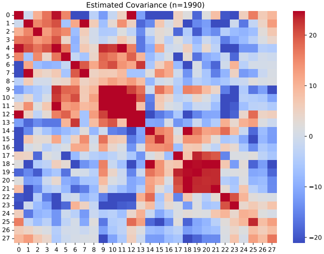
Visualizing the Multivariate Normal Distribution



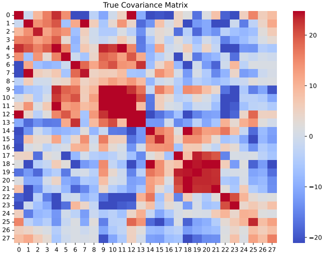
Covariance Estimation in High Dimensions


Estimated Covariance (n=40)

Small sample (n=50)
noisy, unreliable estimate


Estimated Covariance (n=1990)

Large sample (n=2000)
accurate estimate


True Covariance Matrix

True covariance
structure

Flipping the Matrix: Two Views of Data

Let's look at data matrix $\mathbf{X} \in \mathbb{R}^{n \times p}$ (n cells, p genes) from two perspectives:

We can compute two types of covariance matrices:

- Covariance between features (genes):

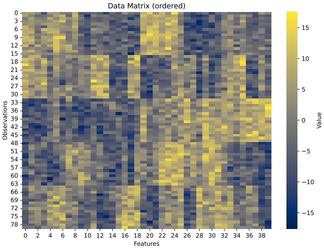
$$\hat{\mathbf{\Sigma}} = \frac{1}{n-1}\mathbf{X}_c^T \mathbf{X}_c$$

- Covariance between observations (cells):

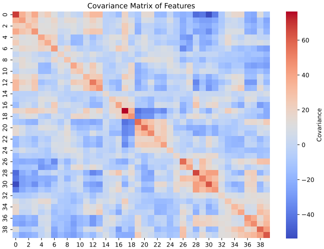
$$\hat{\mathbf{\Sigma}}_{\text{obs}} = \frac{1}{p-1}\mathbf{X}_{c'} \mathbf{X}_{c'}^T$$

Each perspective reveals different aspects of the data structure: how genes relate to each other versus how cells relate to each other.

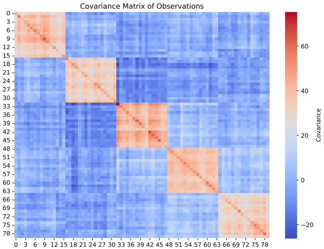
Covariance of Features vs. Observations



Data matrix
(cells × genes)



Feature covariance
(genes × genes)



Observation covariance
(cells × cells)

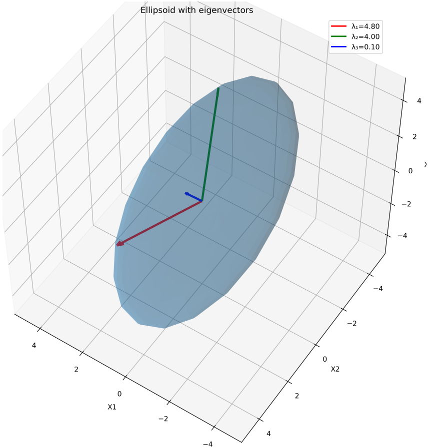
Geometry of the Multivariate Normal

The level sets of a multivariate normal are ellipsoids characterized by the covariance matrix's eigendecomposition:

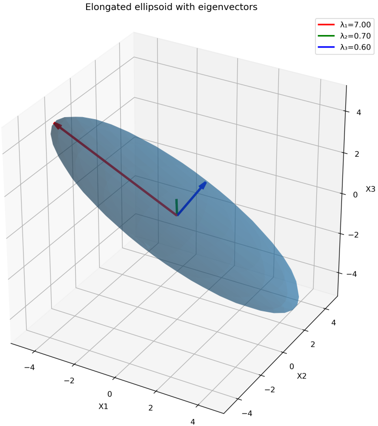
- The eigenvalues λ_i of $\mathbf{\Sigma}$ determine the lengths of the semi-axes of the ellipsoid
- Specifically, the *i*-th semi-axis has length $\sqrt{\lambda_i}$
- The eigenvectors of $\mathbf{\Sigma}$ determine the orientation of the ellipsoid

Sorting eigenvalues in descending order gives the principal axes of variation in the data, from most important to least important.

Ellipsoids with Eigenvectors and Eigenvalues



Eigenvectors (colored arrows) show principal directions;



Elongated ellipsoid with one dominant eigenvector (red).

Interpretation: Principal Components

The eigendecomposition of the covariance matrix provides the principal components of the data:

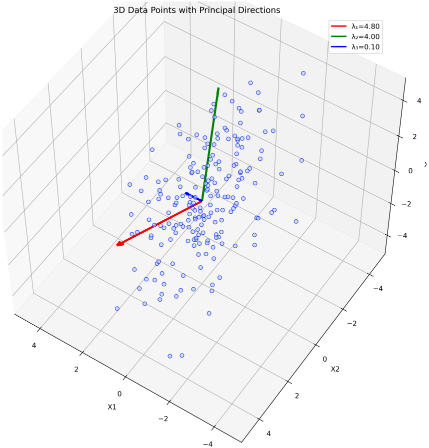
$$\boldsymbol{\Sigma}\mathbf{v}_i = \lambda_i \mathbf{v}_i \quad \text{where} \quad \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$$

- The first principal component $\mathbf{v}_1$ is the direction of maximum variance
- Each subsequent component is orthogonal to previous ones and captures the next most variance
- The eigenvalues λ_i represent the amount of variance explained by each component

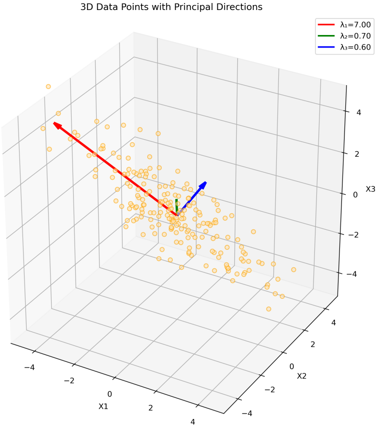
Mathematically, the first principal component is:

$$\mathbf{v}_1 = \underset{\|\mathbf{v}\|=1}{\arg\max}\, \mathrm{Var}(\mathbf{v}^T\mathbf{X})$$

Principal Components in 3D Data



Data with two dominant directions of variation forming a disk-like shape.



Elongated data where one principal component (red arrow) captures most variation.

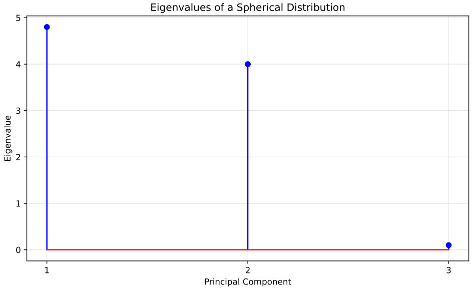
Analysis of Eigenvalues

The eigenvalues of the covariance matrix reveal the underlying structure of the data:

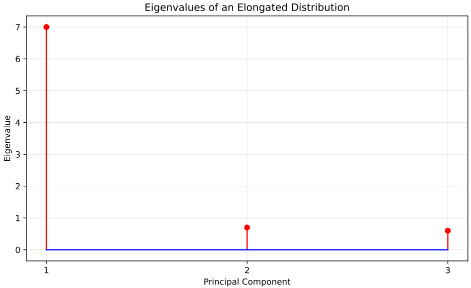
- Eigenvalues represent variance along principal directions
- Sorted eigenvalues ($\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$) show how variance is distributed
- The rank of the covariance matrix equals the number of non-zero eigenvalues
- The distribution of eigenvalues characterizes the "shape" of data in high-dimensional space

By examining eigenvalue patterns, we can determine whether data shows spherical, disk-like, elongated, or more complex structure.

Eigenvalues of Different Covariance Structures

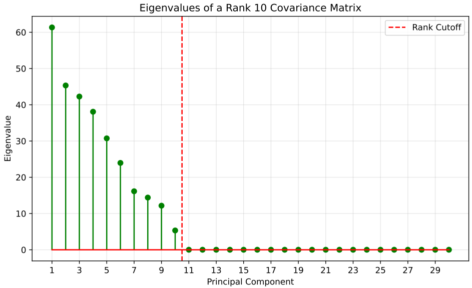


Disk-shaped distribution: two eigenvalues have similar magnitude.

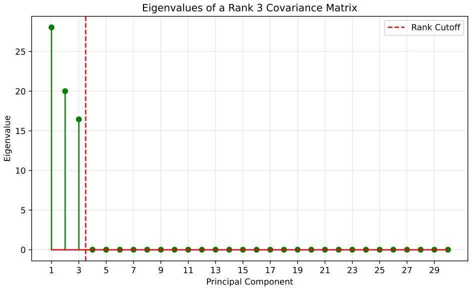


Elongated distribution: one or few eigenvalues are substantially larger, indicating dominant directions of variation.

Eigenvalues of Low-Rank Covariance Matrices



Rank 10 covariance matrix: first 10 eigenvalues are non-zero, suggesting 10 underlying factors explain all variation.



Rank 3 covariance matrix: only 3 non-zero eigenvalues, indicating just 3 latent factors drive all variation.

Principal Component Analysis (PCA)

An Anticipation: Beyond Pairwise Analysis

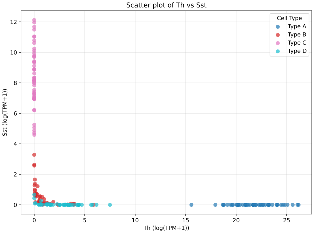
Traditional data exploration often relies on pairwise scatter plots:

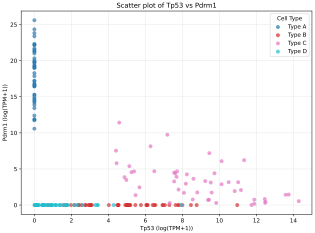
- Variables vs. variables (e.g., gene expression levels)
- Observations vs. observations (e.g., cell similarity)

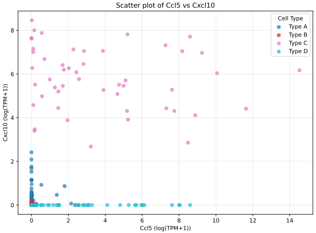
But what if we could get a more global view by transforming the data into a more revealing coordinate system?

Principal Component Analysis transforms high-dimensional data into a coordinate system where axes (principal components) are ordered by importance, revealing global structure that pairwise comparisons might miss.

Pairwise Scatter Plots of Genes

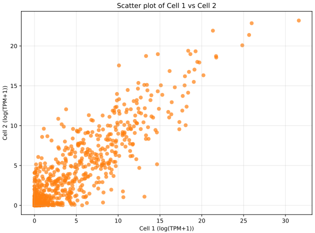


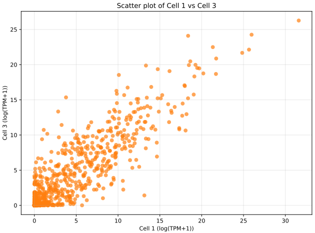


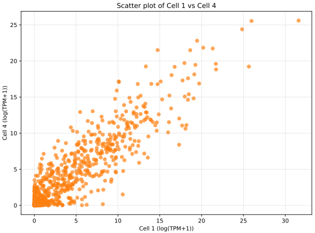


Each pairwise comparison reveals only a fragment of the data's structure. While informative, we're only seeing two dimensions of our high-dimensional data at once, missing the global picture.

Scatter Plots of Cells

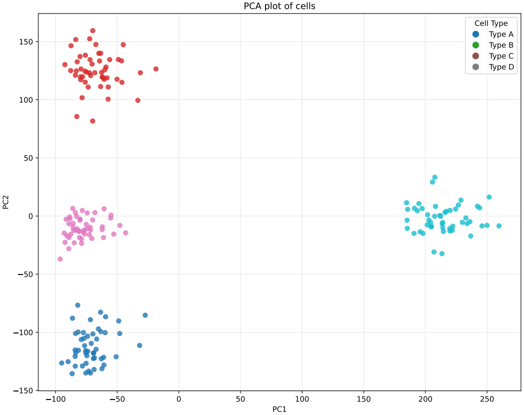

Scatter plot of Cell 1 vs Cell 2


Scatter plot of Cell 1 vs Cell 3


Scatter plot of Cell 1 vs Cell 4

We can also compare pairs of cells, where points represent genes. This provides a different perspective but is still limited to examining two cells at a time.

PCA Provides a Global View



PCA plot of cells, where each point is a cell and axes represent principal components. Notice how cells form distinct clusters that weren't apparent in pairwise gene plots.

PCA Also Works for Variables

PCA plot of genes, where each point is a gene. Genes with similar expression patterns across cells cluster together, revealing functional relationships.

From Eigendecomposition to PCA

While the eigendecomposition of the covariance matrix provides theoretical insight, practical computation of PCA faces challenges:

- Computing and storing the full $p \times p$ covariance matrix can be memory-intensive
- When $p > n$ (more variables than observations), the sample covariance matrix is singular
- Numerical stability issues can arise with ill-conditioned matrices

These practical challenges motivate us to use Singular Value Decomposition (SVD) instead, which works directly with the data matrix rather than the covariance matrix.

Singular Value Decomposition (SVD)

For our centered data matrix $\mathbf{X}_c \in \mathbb{R}^{n \times p}$, the SVD gives us:

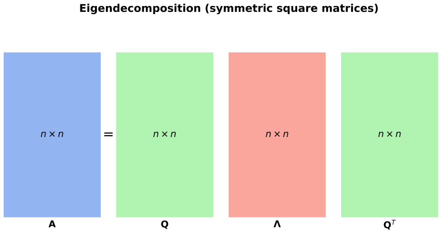
$$\mathbf{X}_c = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$$

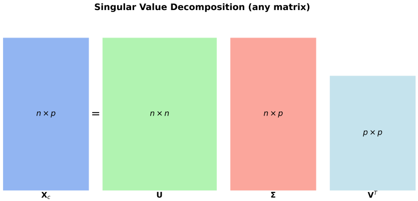
Where:

- $\mathbf{U} \in \mathbb{R}^{n \times n}$ contains the left singular vectors
- $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times p}$ has singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq 0$ on the diagonal
- $\mathbf{V} \in \mathbb{R}^{p \times p}$ contains the right singular vectors

This decomposition works directly with the data matrix, avoiding the need to form the covariance matrix explicitly.

Eigendecomposition vs SVD





Eigendecomposition: A square matrix is decomposed into eigenvectors and eigenvalues. Used for analyzing the covariance matrix.

SVD: Any matrix is decomposed into orthogonal matrices and singular values. Works directly with the data matrix.

Connection Between SVD and the Covariance Matrix

The SVD of the data matrix is directly connected to the eigendecomposition of the covariance matrix:

$$\begin{aligned} \hat{\mathbf{\Sigma}} &= \frac{1}{n-1}\mathbf{X}_c^T\mathbf{X}_c && (1) \\ &= \frac{1}{n-1}\mathbf{V}\mathbf{\Sigma}^T\mathbf{U}^T\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T && (2) \\ &= \mathbf{V}(\frac{1}{n-1}\mathbf{\Sigma}^T\mathbf{\Sigma})\mathbf{V}^T && (3) \end{aligned}$$

This reveals that:

- The right singular vectors **V** are exactly the eigenvectors of the covariance matrix
- The singular values σ_i relate to eigenvalues as $\lambda_i = \sigma_i^2/(n-1)$

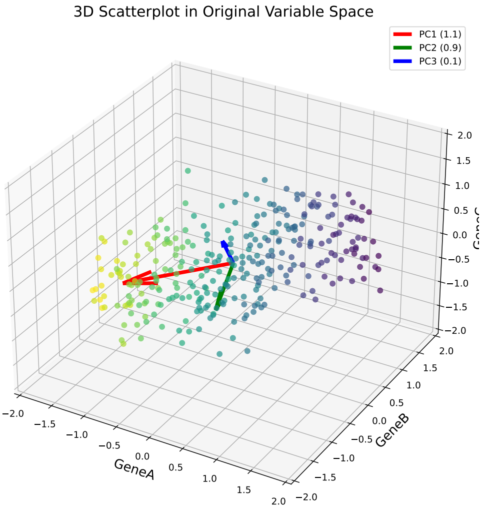
SVD gives us principal components without explicitly forming the covariance matrix.

The Dual Nature of PCA: Variables and Observations

PCA reveals a fascinating duality between the variable space and observation space:

- The right singular vectors **V** (columns of **V**) are eigenvectors of $\mathbf{X}_c^T \mathbf{X}_c$
  - These describe directions in the variable space (e.g., "metagenes")
  - Coefficients are called "loadings": Loading of variable j on PC $k = V_{jk}$
- The left singular vectors **U** (columns of **U**) are eigenvectors of $\mathbf{X}_c \mathbf{X}_c^T$
  - These describe directions in the observation space
  - Coordinates called "PC scores": PC score for observation i on PC $k = U_{ik} \cdot \sigma_k$

PC Coordinates as a Change of Basis

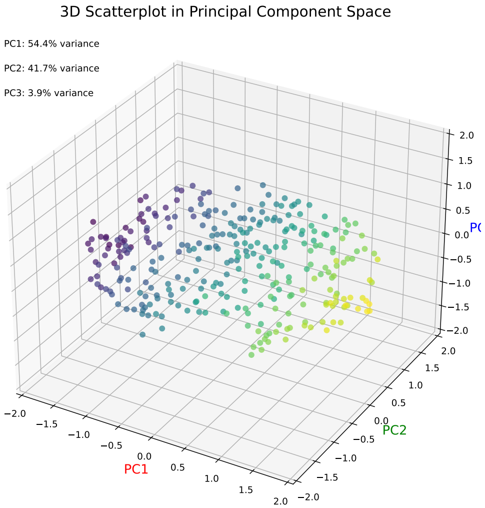


3D scatterplot in original variable space.

Data approximately forms a disk embedded in 3D space.

Red arrows show the principal components.

PC Coordinates as a Change of Basis



The same data viewed in the principal component coordinate system.

This change of basis reveals the intrinsic 2D disk-like structure of the data.

Question 1

You run PCA on a dataset and obtain eigenvalues $\lambda = [11, 7, 6, 3, 2, 1]$. You decide to retain the first three principal components. What proportion of the total variance is explained by these three components?

A) 20%

B) 60%

C) 80%

D) 90%

E) None of the above

Question 2

Based on the relative magnitudes of the eigenvalues $\lambda = [11, 7, 6, 3, 2, 1]$, where the first two are significantly larger than the rest but not exactly equal, what shape best describes the multivariate data cloud?

A) Cigar-shaped
B) Disk-shaped
C) Spherical
D) Rugby-ball-shaped
E) None of the above

Question 3

In PCA, what do the *loadings* represent?

A) The coefficients of the *observations* on each principal component
B) The coefficients of the original *variables* in each principal component
C) The correlations between observations
D) The eigenvalues associated with each principal component
E) The distances of observations from the global mean

Question 4

You calculate the covariance matrix of your data and obtain the following 9×9 block-diagonal matrix (variables 1–3, 4–6, and 7–9 form independent blocks):

$$\Sigma = \begin{pmatrix} 10 & 9 & 8 & 0 & 0 & 0 & 0 & 0 & 0 \\ 10 & 10 & 5 & 0 & 0 & 0 & 0 & 0 & 0 \\ 6 & 10 & 5 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 9 & 2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 11 & 7 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 9 & 9 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & 5 & 4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 7 & 1 & 9 \end{pmatrix}$$

How many blocks of correlated variables would you expect PCA to capture?

A) 1

B) 2

C) 3

D) 4

E) Cannot tell

Question 5

Which of the following statements about PCA is FALSE?

A) The first principal component captures the maximum variance in the data
B) Principal components are orthogonal to each other
C) The number of principal components equals the number of original variables
D) PCA can increase the dimensionality of the data
E) Principal components are linear combinations of the original variables