1 Linear Algebra

1.1 Vector

A vector is an ordered set of numbers that can represent a point in space, direction, or any quantity that has both magnitude and direction.

1.1.1 Inner Product

The inner product (or dot product) of two vectors $\mathbf{u} = [u_1, u_2, \ldots, u_n]$ and $\mathbf{v} = [v_1, v_2, \ldots, v_n]$ is defined as:

$$\mathbf{u} \cdot \mathbf{v} = \sum_{i=1}^{n} u_i v_i.$$

It measures the cosine of the angle between two vectors and their lengths.

1.1.2 Norms

Norms are functions that assign a strictly positive length or size to each vector in a vector space, except for the zero vector. Several types of norms are commonly used:

- **p-norm (Generalized Norm):** The p-norm of a vector $\mathbf{v} = [v_1, v_2, \ldots, v_n]$ is defined as:

  $$\|\mathbf{v}\|_p = \left(\sum_{i=1}^{n} |v_i|^p \right)^{\frac{1}{p}},$$

  where $p \geq 1$. For different values of p, the p-norm takes specific forms:

  - For $p = 1$, it becomes the **L1 norm** (Manhattan norm):

    $$\|\mathbf{v}\|_1 = \sum_{i=1}^{n} |v_i|.$$

  - For $p = 2$, it becomes the **L2 norm** (Euclidean norm):

    $$\|\mathbf{v}\|_2 = \sqrt{\sum_{i=1}^{n} v_i^2}.$$

  - As $p \to \infty$, it approaches the **L∞ norm** (Maximum norm):

    $$\|\mathbf{v}\|_\infty = \max_i |v_i|.$$

1.1.3 Orthogonal and Orthonormal Vectors

Two vectors $\mathbf{u}$ and $\mathbf{v}$ are orthogonal if their inner product is zero:

$$\mathbf{u} \cdot \mathbf{v} = 0.$$

Vectors are orthonormal if they are both orthogonal and have a unit norm ($\|\mathbf{u}\| = 1$).

1.1.4 Linear Independence

A set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is linearly independent if no vector in the set can be written as a linear combination of the others:

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k = 0 \implies c_1 = c_2 = \cdots = c_k = 0.$$

1.2 Matrix

A matrix is a rectangular array of numbers arranged in rows and columns. It is used to represent linear transformations and solve systems of linear equations.

1.2.1 Transpose ($A^\top$)

The transpose of a matrix A is obtained by swapping its rows and columns. If A is an $m \times n$ matrix, its transpose $A^\top$ is an $n \times m$ matrix.

1.2.2 Trace ($\mathbf{tr}(A)$)

The trace of a square matrix A is the sum of its diagonal elements:

$$\operatorname{tr}(A) = \sum_{i=1}^{n} A_{ii}.$$

1.2.3 Inverse (A^{-1})

The inverse of a matrix A is a matrix A^{-1} such that:

$$AA^{-1} = A^{-1}A = I,$$

where I is the identity matrix. A matrix is invertible only if it is square and has full rank.

1.2.4 Orthogonal Matrix

A matrix Q is orthogonal if its transpose is equal to its inverse:

$$Q^\top Q = QQ^\top = I.$$

Orthogonal matrices preserve lengths and angles.

1.3 Matrix Frobenius Norm

The **Frobenius norm** of a matrix $A \in \mathbb{R}^{m\times n}$ is a measure of the magnitude of the matrix elements. It is defined as the square root of the sum of the absolute squares of its elements:

$$\|A\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}|^2},$$

where a_{ij} represents the element in the i-th row and j-th column of the matrix A.

2 Calculus

2.1 Chain Rule

The chain rule is used to differentiate compositions of functions. If $y = f(g(x))$, then the derivative of y with respect to x is:

$$\frac{dy}{dx} = \frac{df}{dg} \cdot \frac{dg}{dx}.$$

2.2 Critical Points

Critical points of a function are points where its derivative is zero or undefined. They are used to find local maxima, minima, or saddle points.

2.3 Taylor Series

The Taylor series of a function $f(x)$ around a point a is given by:

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots .$$

It approximates a function as a polynomial using its derivatives.

3 Probability

Probability theory deals with the analysis of random phenomena and quantifies uncertainty.

3.1 Conditional Probability

The probability of an event A given that another event B has occurred is:

$$P(A|B) = \frac{P(A \cap B)}{P(B)},$$

where $P(A \cap B)$ is the joint probability of A and B.

3.2 Independence and Conditional Independence

Two events A and B are independent if:

$$P(A \cap B) = P(A)P(B).$$

Conditional independence means that two events A and B are independent given a third event C.

3.3 Expectation

The expectation (or mean) of a random variable X is defined as:

$$E[X] = \sum_x xP(X = x),$$

for discrete random variables, or:

$$E[X] = \int_{-\infty}^{\infty} xf(x)\,dx,$$

for continuous random variables.

3.4 Variance and Covariance

The variance of a random variable X measures its spread and is defined as:

$$\mathrm{Var}(X) = E[(X - E[X])^2].$$

Covariance between two random variables X and Y is:

$$\mathrm{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])].$$

3.5 Covariance Matrix

The covariance matrix generalizes the concept of variance to multiple dimensions. For a random vector $\mathbf{X} = [X_1, X_2, \ldots, X_n]^\top$, the covariance matrix Σ is an $n \times n$ matrix where each element Σ_{ij} represents the covariance between X_i and X_j.

$$\Sigma = \begin{bmatrix} \mathrm{Cov}(X_1, X_1) & \mathrm{Cov}(X_1, X_2) & \cdots & \mathrm{Cov}(X_1, X_n) \\ \mathrm{Cov}(X_2, X_1) & \mathrm{Cov}(X_2, X_2) & \cdots & \mathrm{Cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(X_n, X_1) & \mathrm{Cov}(X_n, X_2) & \cdots & \mathrm{Cov}(X_n, X_n) \end{bmatrix}.$$

Each element of the covariance matrix $\Sigma_{ij} = \mathrm{Cov}(X_i, X_j)$ is defined as:

$$\mathrm{Cov}(X_i, X_j) = E[(X_i - E[X_i])(X_j - E[X_j])],$$

Where $E[X_i]$ and $E[X_j]$ are the expected values (means) of the random variables X_i and X_j, respectively.

3.5.1 Example: Covariance Matrix for a 3-Dimensional Dataset

Figure 1 shows a scatter plot matrix of three variables, demonstrating the covariance between them. Positive covariance is indicated by an upward trend, while weaker or zero covariance shows a more scattered pattern.

Consider a random vector $\mathbf{X} = [X_1, X_2, X_3]^\top$ where we have 200 samples of three correlated variables. The computed covariance matrix from the data is:

$$\Sigma = \begin{bmatrix} 1.06 & 0.78 & 0.45 \\ 0.78 & 1.01 & 0.32 \\ 0.45 & 0.32 & 0.98 \end{bmatrix}.$$

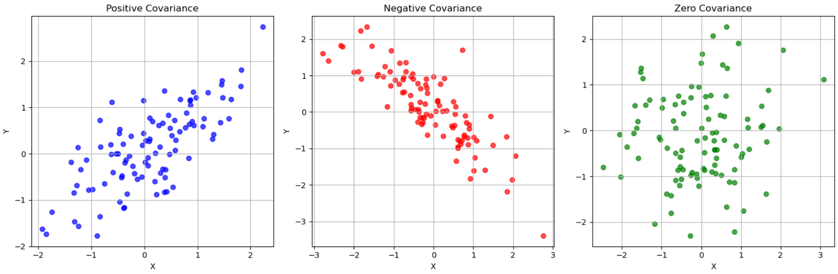

Figure 1: Scatter Plot Matrix for 3-Dimensional Data Showing Covariance

This covariance matrix tells us the following:

1. **Diagonal Elements**: - $\text{Var}(X_1) = 1.06$, $\text{Var}(X_2) = 1.01$, $\text{Var}(X_3) = 0.98$ represent the variances of X_1, X_2, and X_3, respectively.

2. **Off-diagonal Elements**: - $\text{Cov}(X_1, X_2) = 0.78$ indicates a strong positive covariance between X_1 and X_2. - $\text{Cov}(X_1, X_3) = 0.45$ shows a moderate positive covariance between X_1 and X_3. - $\text{Cov}(X_2, X_3) = 0.32$ suggests a weaker positive covariance between X_2 and X_3.

Interpretation of the Covariance Matrix The **positive values** in the off-diagonal elements indicate that as one variable increases, the other variable also tends to increase, suggesting a positive linear relationship. The **magnitude** of the covariance indicates the strength of the relationship: A higher covariance value (e.g., 0.78 between X_1 and X_2) implies a stronger linear relationship compared to a lower value (e.g., 0.32 between X_2 and X_3).