**CIA 1-Component 1 (Assignment)**

**Roll Number: 2448501**

**Name: A ANTO KELVIN PRABHU**

**Assignment Title: CIA1\_Component1**

**Course: Introduction to Statistics for Machine Learning**

**Topic:**

***Assignment Question: Probability Distributions for Continuous Data***

1. ***Normal Distribution***

* ***Describe the probability density function (PDF), key properties, and common applications.***

1. ***Logistic Distribution***

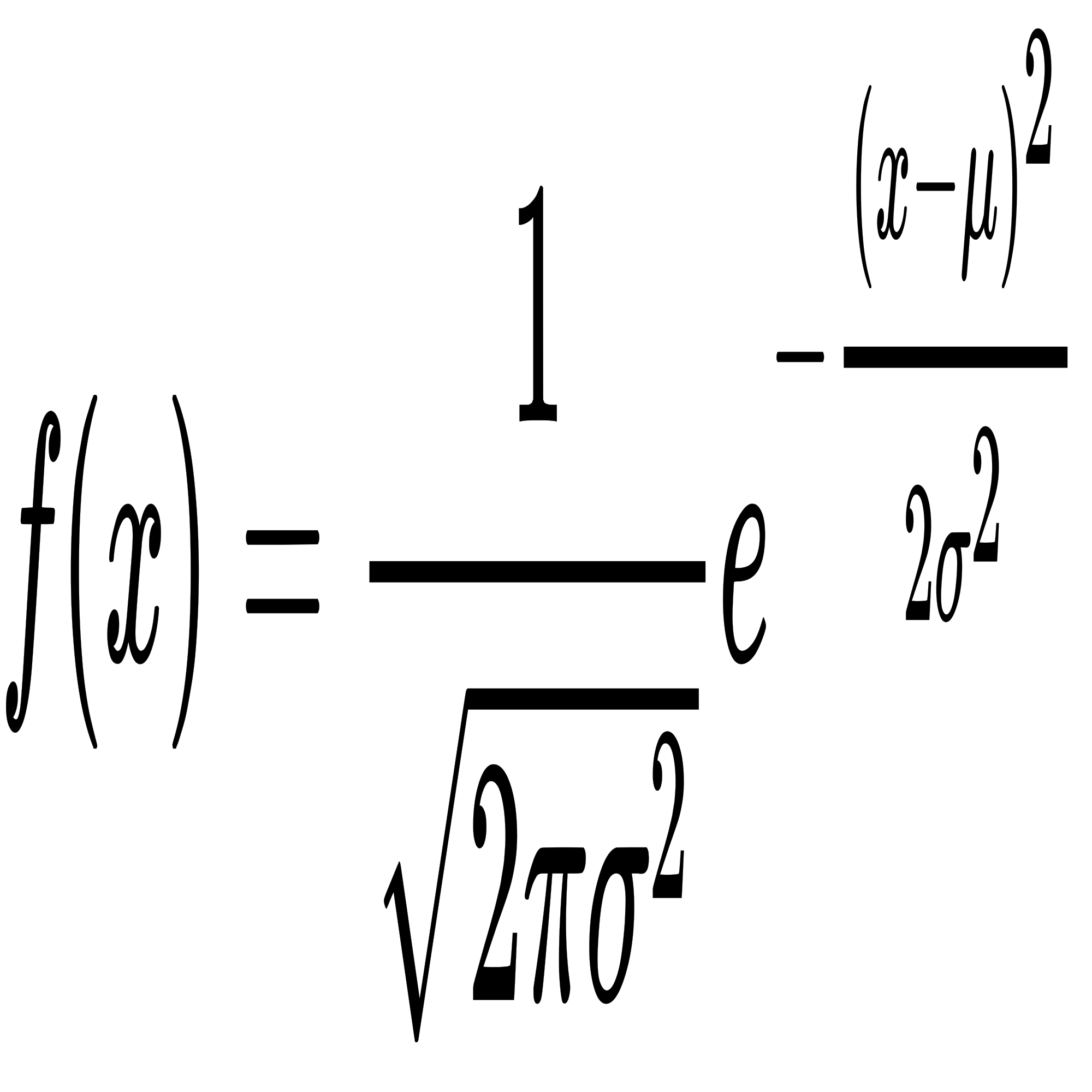
* ***Explain the probability density function (PDF), key properties, and common applications.***

1. ***Multivariate Normal Distribution***

* ***State the probability density function (PDF), mean vector, and covariance matrix.***

**Topic 1: Normal Distribution**

**A normal distribution, often known as a Gaussian distribution, is a kind of continuous probability distribution for a real-valued random variable in probability theory and statistics. Its probability density function's general form is**



**The variance of the distribution is represented by parameter σ 2, whereas the mean or expectation is represented by parameter , along with its median and mode. The distribution's standard deviation is σ sigma . A normal deviate is a random variable that has a Gaussian distribution and is considered normally distributed.**

* 1. **The probability density function (PDF).**

f(x|μ, σ^2) = (1 / √(2πσ^2)) \* exp(- (x - μ)^2 / (2σ^2))

Where:

- μ is the mean,

- σ^2 is the variance.This gets even better if we realize that

1 √ 2 = 1 √ 2π e−1 2z2 2π

is just a constant to make the whole thing integrate to 1. Call that constant

C: f Z(z) = Ce−1 2z2

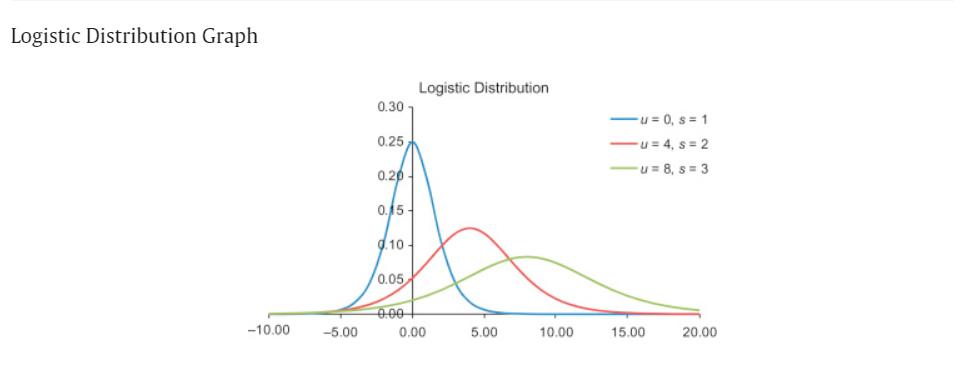
* 1. **Key properties.**

1. Symmetry: With respect to the mean (𝜇μ), the normal distribution is symmetric.
2. Bell-shaped curve: The curve tapers off symmetrically on both sides, with the mean serving as the highest point.
3. Mode = Mean - Median All three of these measures of central tendency are equal in a normal distribution.
4. The standard deviation, or σ σ, establishes the spread. When σ rises, the curve widens.
5. 68, 95, 99.7 Rule: Approximately 68% of the data are within one standard deviation, 95% are within two, and 99.7% are within three of the mean.
   1. **Common applications.**

Numerous industries, including quality assurance, finance (for stock returns), and the study of natural phenomena like human traits, use the normal distribution (e.g., height, weight)

**Topic 2: Logistic Distribution**

**A continuous distribution function is the logistic distribution. Its pdf and cdf functions are employed in a wide range of applications, including neural networks, logistic regression, and logit models. It has been applied to modelling in sports, the physical sciences, and most recently, finance. Because the logistic distribution's tails are larger than those of a normal distribution, it better captures the chance of extreme events and is therefore more consistent with the underlying data.**



**2.1 probability density function (PDF).**

The Probability Density Function (PDF) of the logistic distribution is given by:

f(x|μ, s) = e^{-(x-μ)/s} / [s(1 + e^{-(x-μ)/s})^2]

Where:

- μ is the location parameter (mean),

- s is the scale parameter (similar to standard deviation).

**2.2 Key properties.**

- Symmetry: The logistic distribution is symmetric about its mean, just as the normal distribution.

- Faster Tails: The logistic distribution predicts extreme values more frequently than the normal distribution because it has heavier tails.

- The logistic (S-shaped) curve of the cumulative distribution function (also known as the CDF) is helpful in modelling growth.

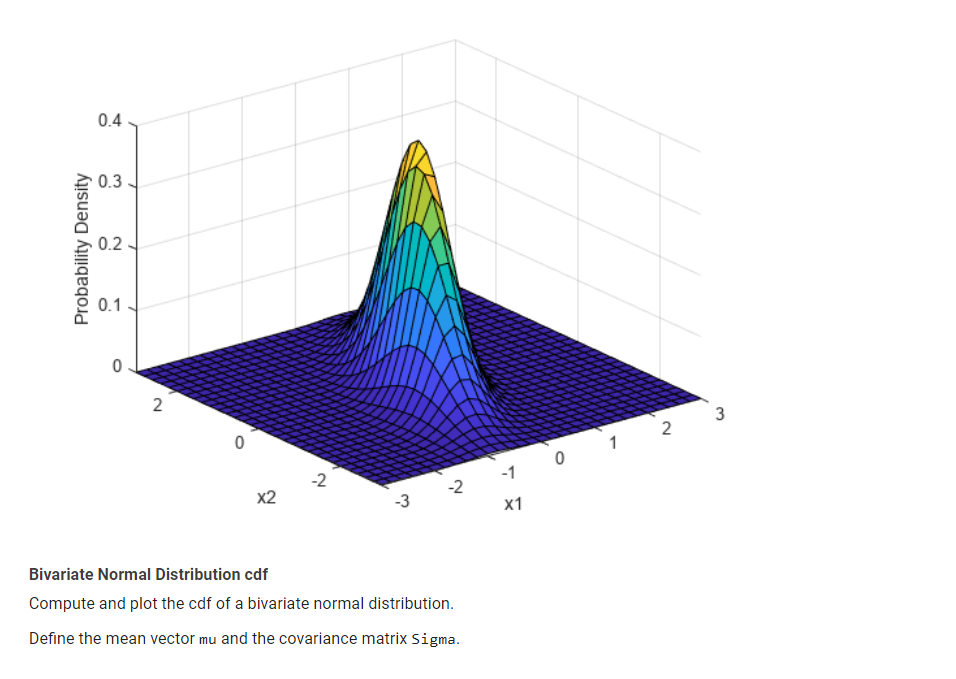
**2.3 Common applications.**

In biological research, the logistic distribution is frequently employed as a model for population growth and in logistic regression and survival analysis. Additionally, machine learning—more especially, binary classification models—applies it.

**Topic 3: Multivariate Normal Distribution**

**An extension of the univariate normal distribution to two or more variables is the multivariate normal distribution. It is a distribution for random vectors of correlated variables with a univariate normal distribution for each vector element. In the most basic scenario, the vector elements are independent univariate normal random variables and there is no correlation between the variables.**

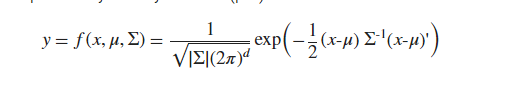
**The multivariate normal distribution is a popular choice for a multivariate data model due to its ease of usage.**



**Numerous features pertaining to the multivariate normal distribution are available through the Statistics and Machine Learning ToolboxTM.**

**3.1 probability density function (PDF).**

The Probability Density Function (PDF) for a multivariate normal distribution with a mean vector μ and covariance matrix Σ is:

where x and μ are 1-by-d vectors and Σ is a d-by-d symmetric, positive definite matrix.

Statistics and Machine Learning Toolbox:

Supports singular Σ for random vector generation only. The pdf cannot be written in the same form when Σ is singular.

Uses x and μ oriented as row vectors rather than column vectors

Where:

- μ is the mean vector,

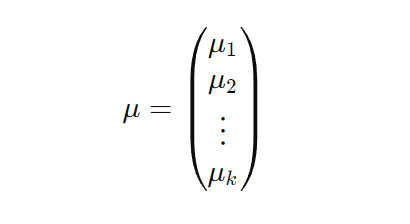
- Σ is the covariance matrix,

- k is the number of variables (dimensions),

- x is a vector of random variables.

**3.2 mean vector.**

μ in a multivariate normal distribution is a



k-dimensional column vector that represents the expected values (means) of each random variable in the distribution. It is expressed as:

**3.3 Covariance matrix**

The variances along the diagonal and the covariances between various variables off the diagonal are contained in the covariance matrix Σ Σ. It is set up as follows:

The variance of the i-th variable is represented by σ i 2​, while the covariance between the i-th and j-th variables is represented by σ ij​.

The multivariate normal distribution is a higher-dimensional generalisation of the normal distribution, in which the covariance matrix governs the relationships between the variables and each variable has a normal distribution.

**3.4 Application**

1. Used in portfolio theory to model the returns of multiple assets,
2. A key component in multivariate statistical methods such as Principal Component Analysis (PCA),
3. Commonly applied in Gaussian Mixture Models (GMM) in machine learning.

**4. Conclusion**

Continuous probability distributions such as the multivariate normal, logistic, and normal distributions are essential in machine learning and statistical analysis. Because it can be applied in a wide range of real-world scenarios and has symmetry thanks to the central limit theorem, the normal distribution is especially well-liked. Because of its heavier tails, the logistic distribution is frequently employed in binary classification tasks and growth models. The multivariate normal distribution allows us to model relationships between multiple variables, which is an advancement over the normal distribution. Because of their distinctive qualities, each of these distributions can be applied to a variety of situations, ranging from straightforward statistical tests to intricate multivariate analyses.

**References:**

# **YouTube:** Complete Statistics For Data Science In 6 hours By Krish Naik .

**Mathworks:https://www.mathworks.com/help/stats/multivariate-normal-distribution.html**

**Sciencedirect:https://www.sciencedirect.com/topics/mathematics/logistic-distribution**

**Wikipedia:https://en.wikipedia.org/wiki/Normal\_distribution**