Submitted by: Adithya Acharya
USN : 01SU24CS011
Branch: CSE
Section : A section
Subject: Fundamentals of AI and ML
Topic : Assignment 2
Submitted to : Prof. Mahesh Kumar

ASSIGNMENT (Question -2)

Take one Domain and draw the graph (Normal distribution) (Empirical rule)

Introduction

Temperature is one of the most important environmental variables that affects daily life, agriculture, health, and economic activities. In most regions, daily temperatures do not change randomly but follow a predictable pattern around an average value. When observed over a long period, these temperature values often form a symmetrical, bell-shaped pattern known as a normal distribution.

The normal distribution helps us understand how temperature values are spread around the mean temperature. Most days usually experience temperatures close to the average, while extremely hot or extremely cold days occur less frequently. This behavior can be explained using the Empirical Rule (68–95–99.7 rule), which describes how data is distributed within standard deviations from the mean.

Domain Selected: Daily Temperature in a City

In this section, we will analyze daily summer temperature in a city using the concept of normal distribution and apply the empirical rule to understand how temperature values are distributed.

Let us consider the daily average temperature of a city during summer. Temperature values usually fluctuate around a central value and follow a pattern that is approximately normal.

Assume:

Mean temperature (μ) = 30°C

Standard Deviation (σ) = 5°C

This means most days the temperature is close to 30°C, but some days may be slightly cooler or hotter.

What is Normal Distribution?

Normal distribution is a continuous probability distribution that forms a symmetric bell-shaped curve. In this distribution, the majority of values are concentrated around the mean, and the frequency decreases as we move away from the center.

It is also known as Gaussian distribution. Many natural phenomena such as temperature, height, weight, and exam scores follow this pattern because they are influenced by many small independent factors.

Normal Distribution (Daily Life Temperature Domain)

Normal distribution is a statistical concept that describes how data values are spread around an average. In daily life, temperature readings of a city over a long period often follow a normal distribution pattern. Most days have temperatures close to the average, while extremely hot or cold days are less frequent. This creates a bell-shaped curve when the data is plotted on a graph.

Key Points:

- Daily temperatures tend to cluster around an average value.
- Extremely high or low temperatures occur rarely.
- The distribution forms a symmetrical bell-shaped curve.
- It helps in understanding weather patterns and climate trends.

Mean

The mean is the average of all daily temperature readings over a specific period. In a normal distribution, the mean lies at the center of the curve and divides it into two equal halves. It represents the most typical temperature experienced.

Key Points:

- Calculated by adding all temperature values and dividing by total days.
- Located at the center of the distribution.
- Equal to median and mode in a perfect normal distribution.

Standard Deviation

Standard deviation measures how much the daily temperatures vary from the average temperature. A small standard deviation means temperatures remain stable, while a large one indicates greater fluctuations. It determines the width of the bell-shaped curve**.**

Key Points:

- Shows spread of temperature data.
- Small value = less variation.
- Large value = more variation.
- Affects the shape of the distribution curve.

Characteristics of Normal Distribution

A normal distribution is symmetrical and bell-shaped. The total area under the curve equals 1 (or 100%). The mean, median, and mode are equal and located at the center. The curve approaches but never touches the horizontal axis.

Key Points:

- Symmetrical around the mean.
- Bell-shaped curve.
- Mean = Median = Mode.
- Total probability equals 100%.

Empirical Rule

The empirical rule, also known as the 68–95–99.7 rule, explains how data is distributed in a normal distribution. It states that about 68% of values fall within one standard deviation from the mean, 95% within two, and 99.7% within three. This rule helps estimate temperature ranges easily.

Key Points:

- 68% within ±1 standard deviation.
- 95% within ±2 standard deviations.
- 99.7% within ±3 standard deviations.
- Useful for predicting temperature ranges.

1.68% Rule (One Standard Deviation)

About 68% of the data values fall within one standard deviation from the mean ($\mu \pm 1\sigma$). This means that more than half of the individuals or observations are concentrated close to the average value.

Example:

• Height range between 160 cm and 180 cm.

• Represents the most common height range in the population.

This region forms the central portion of the distribution where most values are located.

2.95% Rule (Two Standard Deviations)

Approximately 95% of the data values lie within two standard deviations from the mean ($\mu \pm 2\sigma$). This means that nearly all individuals or observations are included within this wider range around the average.

Example:

• Height range between 150 cm and 190 cm.

• Covers most of the population, excluding only a small number of extreme values.

This interval represents a large majority of the distribution.

3.99.7% Rule (Three Standard Deviations)

Almost 99.7% of the observations fall within three standard deviations from the mean ($\mu \pm 3\sigma$). This indicates that nearly every value in the dataset is contained within this range.

Example:

• Height range between 140 cm and 200 cm.

• Only a very tiny fraction of individuals fall outside this boundary.

This range practically includes the entire distribution.

Example in Daily Temperature

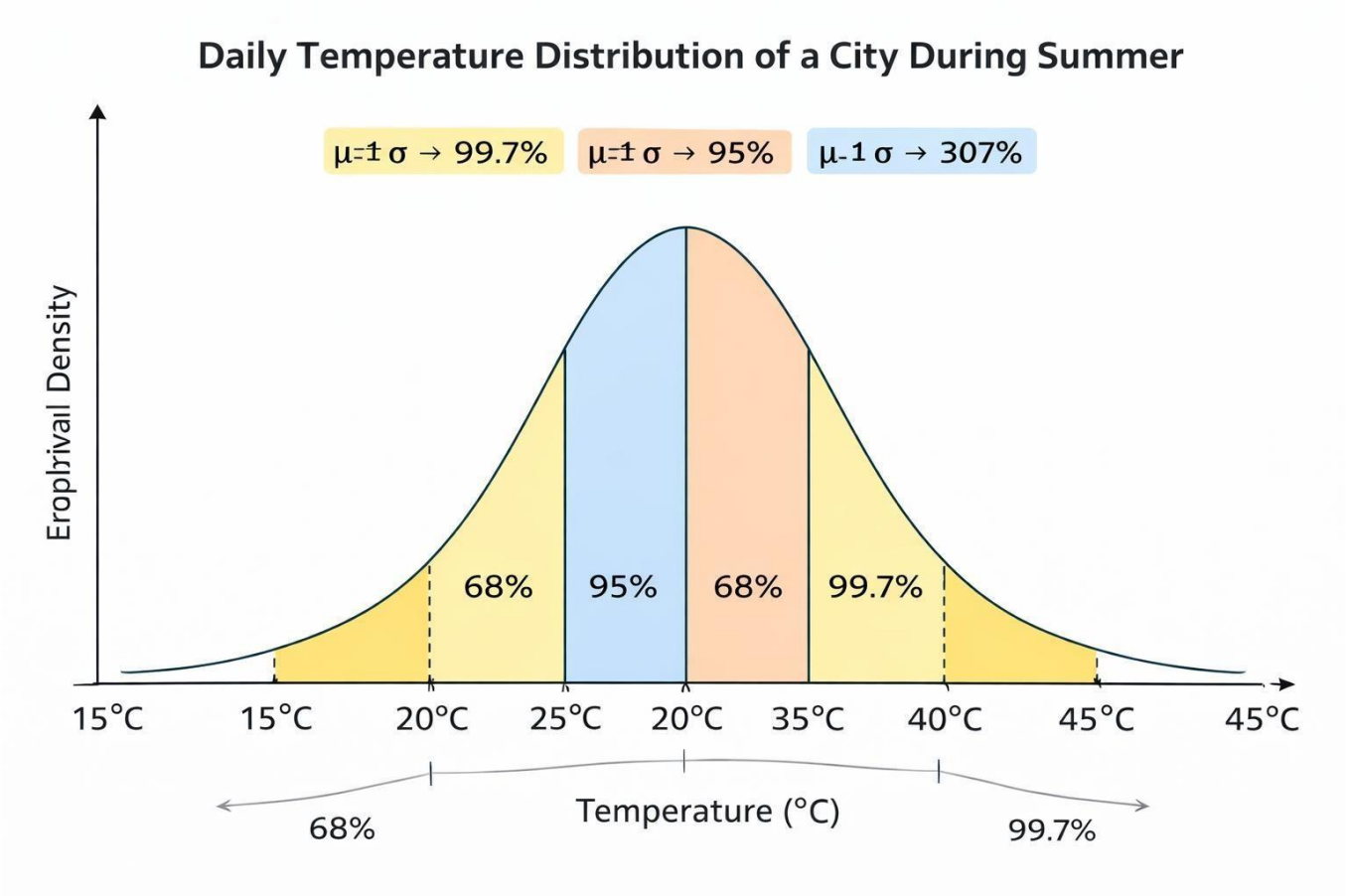
Suppose the average daily temperature of a city is 30°C with a standard deviation of 2°C. Around 68% of the days, the temperature will range between 28°C and 32°C. About 95% of the days will fall between 26°C and 34°C. Very few days will have temperatures below 24°C or above 36°C.

Key Points:

- Mean temperature = 30°C.
- Standard deviation = 2°C.
- Most days fall between 28°C–32°C.
- Extreme temperatures are rare.

This forms the central region of the bell curve.

Graph Representation of Human Height Distribution



The normal distribution graph of daily temperature is represented by a bell-shaped curve.

Components of the Graph

X-axis

Represents data values (e.g., temperature, height, marks).

Y-axis

Represents frequency or probability.

Mean (μ)

The mean is the average value of all observations in a dataset. In a bell-shaped graph, the mean is located at the center and divides the curve into two equal halves.

Standard Deviations (Daily Temperature Example)

Marked as:

• $\mu - 1\sigma$, $\mu + 1\sigma \rightarrow$ Normal daily temperature range (most common temperatures)

• $\mu - 2\sigma$, $\mu + 2\sigma \rightarrow$ Slightly higher and lower than average temperature range

• $\mu - 3\sigma$, $\mu + 3\sigma \rightarrow$ Extremely hot or extremely cold temperature range

The curve remains symmetric on both sides of the mean temperature, showing that temperature variations above and below the average occur in a balanced pattern.

Interpretation of Daily Temperature Distribution

Using the empirical rule, daily temperature distribution can be categorized as:

Low Temperature (Below -2σ)

• Very few days

• Much colder than the average temperature

Normal Temperature (Between -1σ and $+1\sigma$)

• Majority of days

• Usual or comfortable temperature range

High Temperature (Above $+1\sigma$)

• Fewer days

• Warmer than average temperature

Extreme Temperature (Beyond $\pm 3\sigma$)

• Very rare days

• Extremely hot or extremely cold conditions

This classification helps in weather analysis, climate studies, and temperature forecasting.

Applications of Normal Distribution in Daily Temperature and Climate Studies

Normal distribution is widely used in weather analysis and climate research.

1. Weather Monitoring

Helps track daily temperature changes over time.

2. Climate Assessment

Used to study long-term temperature patterns and seasonal variations.

3. Environmental Research

Helps analyze temperature trends related to global warming.

4. Regional Climate Analysis

Helps understand temperature characteristics of different cities or regions.

5. Forecasting and Predictions

Used for predicting future temperature trends and weather conditions.

Advantages of Using Normal Distribution

• Represents real-world temperature data effectively.

• Helps in accurate prediction and analysis.

• Easy to interpret temperature trends.

• Useful in meteorology and climate research.

Limitations of Normal Distribution

• Not all temperature data follows a perfect normal pattern.

• Sudden climate changes may create extreme values.

• Requires sufficient data for accurate results.

Despite these limitations, it remains an important statistical tool in weather and climate analysis.

Conclusion

The normal distribution provides an effective method for analyzing real-world data such as daily temperature. The bell-shaped curve shows that most days have temperatures close to the average, while only a few days experience extreme heat or cold. The empirical rule explains how temperatures are distributed within one, two, and three standard deviations from the mean.

Studying temperature distribution helps meteorologists understand climate behavior and seasonal variations. The empirical rule allows better interpretation of temperature data and supports forecasting and environmental planning. Therefore, normal distribution plays a significant role in statistical analysis and real-world temperature studies.