SNS COLLEGE OF ENGINEERING

Kurumbapalayam (Po), Coimbatore – 641 107

An Autonomous Institution

Accredited by NBA – AICTE and Accredited by NAAC – UGC with 'A' Grade
Approved by AICTE, New Delhi & Affiliated to Anna University, Chennai

DEPARTMENT OF COMPUTER SCIENCE AND DESIGNING

COURSE NAME : 19CS401 DATA ANALYTICS

III YEAR /V SEMESTER

Unit 3- CLUSTERING

Topic :Distance Measures in Attributes

What are Data Attributes?

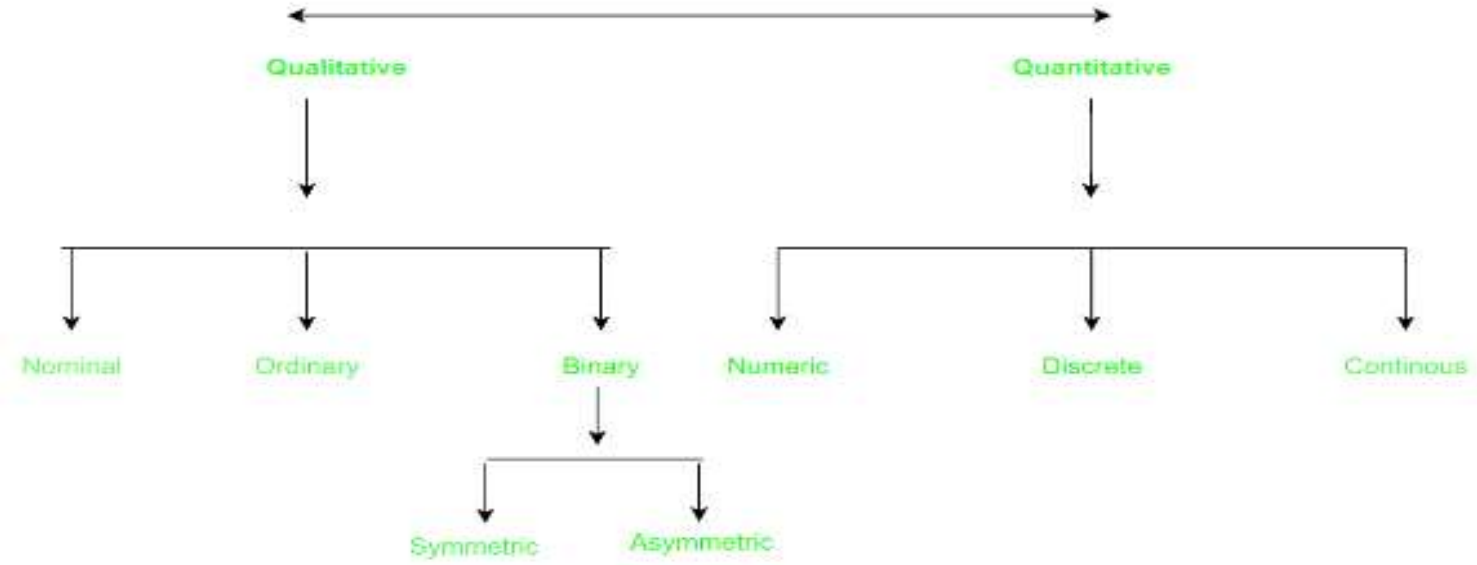
- Attributes are qualities or characteristics that describe an object, individual, or phenomenon.
- Attributes can be categorical, representing distinct categories or classes, such as colors, types, or labels.
- Some attributes are quantitative, taking on numerical values that can be measured or counted, such as height, weight, or temperature.
- Categorical attributes can be further classified as nominal (no inherent order) or ordinal (possessing a meaningful order).
- In data analysis, attributes serve as the variables of interest, and understanding their nature is crucial for selecting appropriate statistical methods and drawing meaningful conclusions from the data.

Types of attributes

This is the initial phase of data preprocessing involves categorizing attributes into different types, which serves as a foundation for subsequent data processing steps. Attributes can be broadly classified into two main types:

1.Qualitative (Nominal (N), Ordinal (O), Binary(B)).

2.Quantitative (Numeric, Discrete, Continuous)



Qualitative Attributes

1. Nominal Attributes :

Nominal attributes, as related to names, refer to categorical data where the values represent different categories or labels without any inherent order or ranking. These attributes are often used to represent names or labels associated with objects, entities, or concepts.

Example :

Attribute	Values
Colours	Black, Brown, White
Categorical Data	Lecturer, Professor, Assistant Professor

Qualitative Attributes

2. Binary Attributes: Binary attributes are a type of qualitative attribute where the data can take on only two distinct values or states. These attributes are often used to represent yes/no, presence/absence, or true/false conditions within a dataset. They are particularly useful for representing categorical data where there are only two possible outcomes. For instance, in a medical study, a binary attribute could represent whether a patient is affected or unaffected by a particular condition.

•**Symmetric:** In a symmetric attribute, both values or states are considered equally important or interchangeable. For example, in the attribute “Gender” with values “Male” and “Female,” neither value holds precedence over the other, and they are considered equally significant for analysis purposes.

Attribute	Values
Gender	Male , Female

Asymmetric: An asymmetric attribute indicates that the two values or states are not equally important or interchangeable. For instance, in the attribute “Result” with values “Pass” and “Fail,” the states are not of equal importance; passing may hold greater significance than failing in certain contexts, such as academic grading or certification exams

Attribute	Values
Cancer detected	Yes, No
result	Pass , Fail

Qualitative Attributes

3. Ordinal Attributes : Ordinal attributes are a type of qualitative attribute where the values possess a meaningful order or ranking, but the magnitude between values is not precisely quantified. In other words, while the order of values indicates their relative importance or precedence, the numerical difference between them is not standardized or known.

Example:

Attribute	Value
Grade	A,B,C,D,E,F
Basic pay scale	16,17,18

Quantitative Attributes

1. Numeric: A numeric attribute is quantitative because, it is a measurable quantity, represented in integer or real values. Numerical attributes are of 2 types: **interval** , and **ratio-scaled.**

•An **interval-scaled** attribute has values, whose differences are interpretable, but the numerical attributes do not have the correct reference point, or we can call zero points. Data can be added and subtracted at an interval scale but can not be multiplied or divided. Consider an example of temperature in degrees Centigrade. If a day's temperature of one day is twice of the other day we cannot say that one day is twice as hot as another day.

•A **ratio-scaled** attribute is a numeric attribute with a fix zero-point. If a measurement is ratio-scaled, we can say of a value as being a multiple (or ratio) of another value. The values are ordered, and we can also compute the difference between values, and the mean, median, mode, Quantile-range, and Five number summary can be given.

2. Discrete : Discrete data refer to information that can take on specific, separate values rather than a continuous range. These values are often distinct and separate from one another, and they can be either numerical or categorical in nature.

Attribute	Value
Profession	Teacher, Business man, Peon
ZIP Code	301701, 110040

Quantitative Attributes

3. Continuous : Continuous data, unlike discrete data, can take on an infinite number of possible values within a given range. It is characterized by being able to assume any value within a specified interval, often including fractional or decimal values.

Example :

Attribute	Value
Height	5.4, 6.2 ...etc
weight	50.33 .....etc

What is a target attribute?

A target attribute, also known as a target variable or response variable, is a specific attribute or column in a dataset that represents the outcome or prediction target in a supervised learning problem. In supervised learning, the goal is typically to predict or model the value of the target attribute based on the values of other attributes, known as predictor variables or features.

For example, in a dataset of housing prices, the target attribute might be the sale price of houses, while the predictor variables could include attributes such as the number of bedrooms, the square footage, and the location. The target attribute is what the model aims to predict or estimate based on the input features.

FAQ's

1.What are data types of attributes?

2. What is the difference between nominal and ordinal attributes?

3. How do discrete and continuous attributes differ?

4. What are attributes in warehouse?


SNS COLLEGE OF ENGINEERING
COIMBATORE -107


sns
INSTITUTIONS

Thank You!