7.3 Attribute Types

Attribute Types

Before we get into the algorithms of machine learning it would be conducive to our learning objective that we get familiar with various attribute types a data set can have.

A collection of rows which we get in a data file is called data set. Basically, it is the data which we work on. Each row in the data file represent one instance of the data and is known as data point they are also known as a data object. Columns in the data set represent attribute or feature of the data.

These attributes can be divided into two broad categories.

1. Discrete attributes: They can only take particular values but those values can be infinite in a given interval. They can be integers, or categorical: male/female or good/bad.
2. Continuous attributes: They can have any value they are represented by floating point integer, e.g., temperature which can be 21.1, 21.11, 63.0, 63.23.

Another classification can be based on mathematical operators like which arithmetical or logical operators can be applied to these attributes.

1. **Nominal**: They are different names using which we can distinguish one from another. This distinctness can be tested using = and ≠ symbol, e.g., Street number, employee ID, Zip code, ... etc.
2. **Ordinal**: This attribute gives a sense of ordering like good, better, best or grade. The ordering can be tested using ≥ or ≤ operator. They have both distinctness and order.
3. **Interval**: They are associated with a unit of measurement, as a result, we can add them or subtract them. + and – is associated with them, e.g., temperature, speed, count, ... etc. We can have distinctness, order, and addition or subtraction operation on them.
4. **Ratio**: We can apply both difference and ratio in a meaningful way, i.e., * and / example length, time, count. We can have all the four operations on them, i.e., distinctness, order, addition or subtraction, and multiplication or division.