SNS COLLEGE OF TECHNOLOGY

Coimbatore-37.
An Autonomous Institution

COURSE NAME : 19CSE301-INTRODUCTION TO DATA SCIENCE

III YEAR/ V SEMESTER

UNIT – I DATA SCIENCE LIFE CYCLE

Topic: Rectangular Data

Ms.G.Swathi
Assistant Professor
Department of Computer Science and Engineering

Rectangular data

- Rectangular data are the staple of statistical and machine learning models. Rectangular data are **multivariate cross-sectional data (i.e. not time-series or repeated measure) in which each column is a variable (feature), and each row is a case or record**
- Rectangular data structures generally store data in a *two-dimensional* (2D) format (i.e., a grid containing rows and columns). When all rows and all columns have the same length, the resulting structure is *rectangular*.

Terms for Rectangular data

Key Terms for Rectangular Data

Data frame

Rectangular data (like a spreadsheet) is the basic data structure for statistical and machine learning models.

Feature

A column within a table is commonly referred to as a *feature*.

Synonyms

attribute, input, predictor, variable

Outcome

Many data science projects involve predicting an *outcome*—often a yes/no outcome .

The *features* are sometimes used to predict the *outcome* in an experiment or a study.

Synonyms

dependent variable, response, target, output

Records

A row within a table is commonly referred to as a *record*.

Synonyms

case, example, instance, observation, pattern, sample

Types of Rectangular Data

Rectangular data are shaped like a rectangle where every value corresponds to some row and column. Most data frames store rectangular data.

Non-rectangular data, on the other hand, are not neatly arranged in rows and columns. Instead, they are often a culmination of separate data structures where there is some similarity among members of the same data structure.

References

- Tom M. Mitchell, “Machine Learning”, McGraw-Hill Education (India) Private Limited, 2013.


Thank
You