Lecture: 03 & 04

(understanding Data)

➡ Types of Data

- ⊙ structured data (spreadsheets, tables)
- ⊙ unstructured data (speech, image, text)

◆ Attributes are also known as Features, columns, variables

→ Attribute is a CS or Database term
→ Feature is a Machine Learning term
→ variable is a statistical term

◆ Data objects are also known as rows, sample, observation, record, instance, Data point

→ Sample is a statistical term
→ observation is a data analysis term
→ Record is a database term
→ Instance is a machine learning term

Types of Attributes

① Nominal (labels or names) (no inherent order) (categorical) (can represented by symbols)
② Binary (specialized form of nominal with only two values) 0 (negative) 1 (positive)
③ Ordinal (order or ranking) can be represted by symbols or codes
④ Numeric (Quantitative attribute that are measurable)

➡ Arithmatic operators are not meaningful on categorical variables

⇒ Binary attribute

Symmetric	Asymmetric
(Both values are equally important)	(one value is more important)
(e.g) Male / Female	(e.g) 1 = smoker , 0 = Non-smoker

➡ Mode and Median are valid for ordinal data

➡ Mode is valid for nominal data as well

➡ Discretization (converting a numeric attribute into ordinal attribute)

numeric attribute

Interval scale attribute	Ratio scale attribute
① No absolute zero	① has meaningful absolute zero point (complete absense)
(e.g) no 'zero' date	(eg) Years of experience is absolute zero
② Measured on equal size scale	② Ratio's b/w values are meaningful

Discrete	VS	Continuous attributes
→ whole numbers (int)		→ floating values
(e.g) no of customers		(e.g) Temp, height, Time duration of process