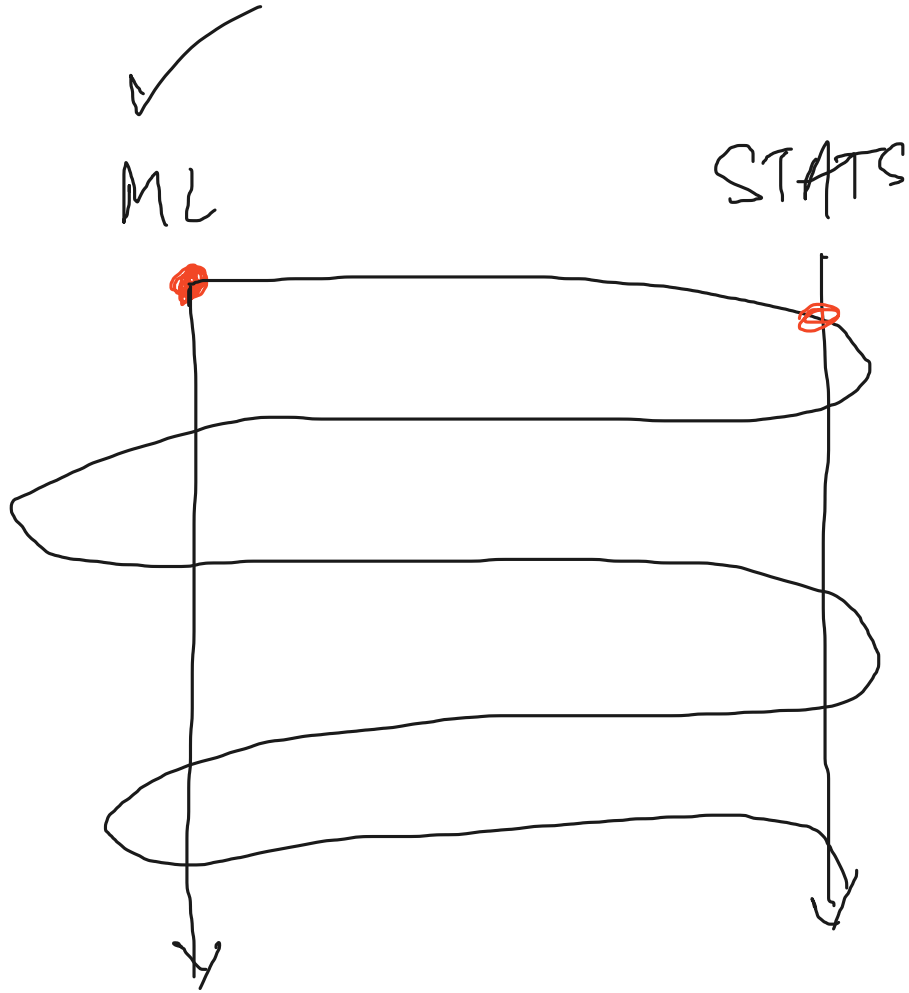
https://tinyurl.com/ds203-2024-q

$$Y = f(X)$$

UNKNOWN.

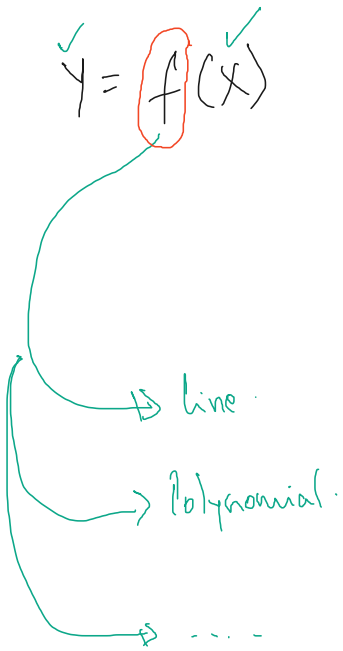
What is ML? → Patterns / Trends.

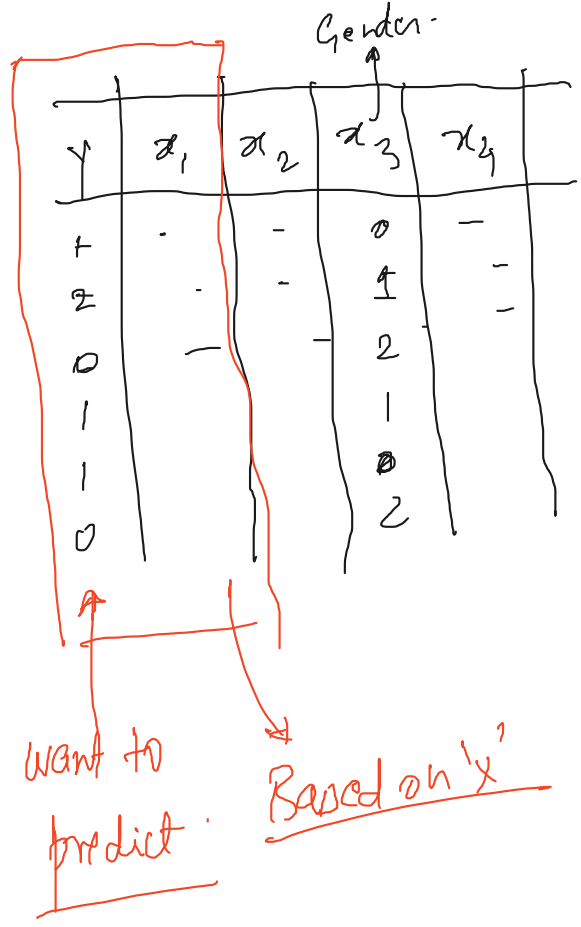


What do we have?

— DATA. (STRUCTURED DATA)

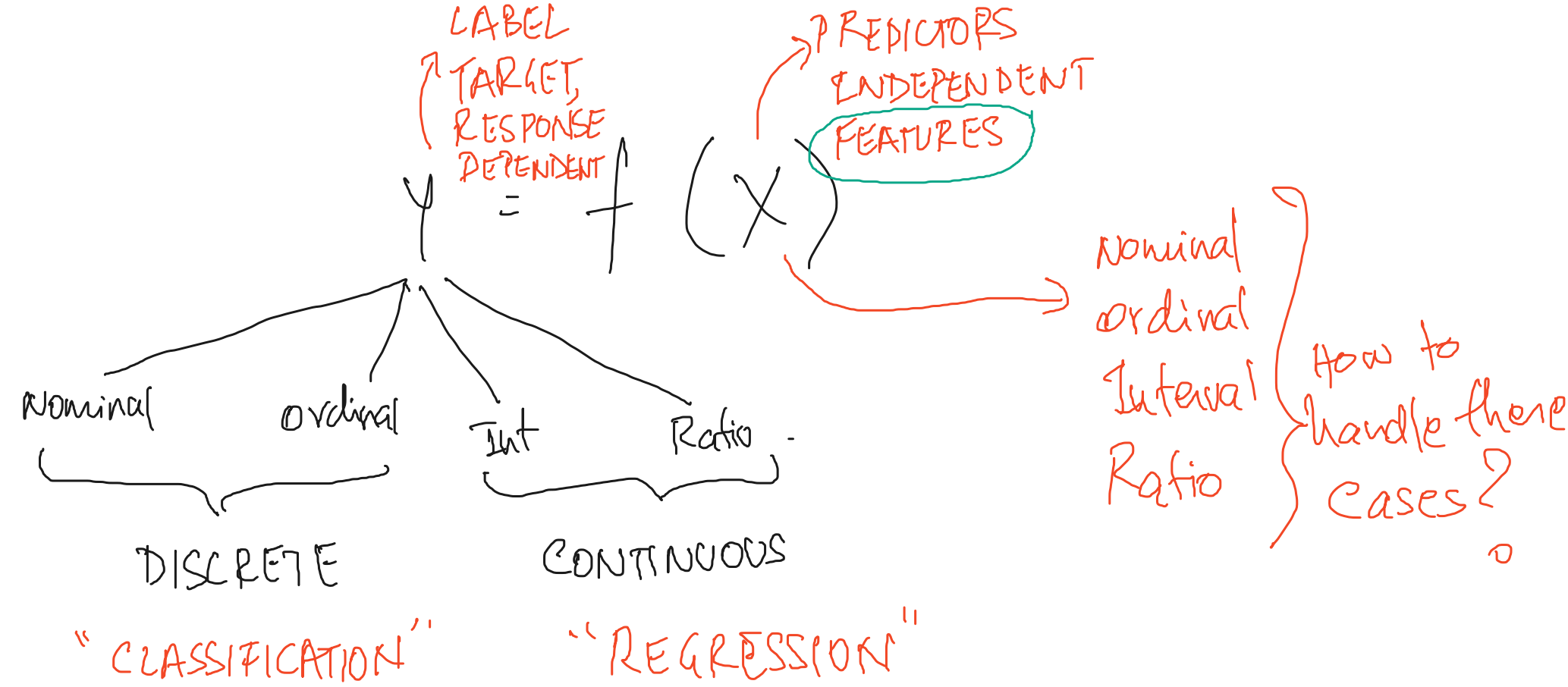
"TABLE"





Levels of Measurement

- NOMINAL .... DISCRETE
  NO ORDER
  DIFFERENTIATE
  Gender — M / F
  Red, Green, Blue
- ORDINAL ... DISCRETE
  ORDER
- INTERVAL - CONTINUOUS
  Differences
  "ZERO" is an arbitrary decision.
  → Temperature
  20°C, 0°, 10°C
- RATIO
  → Continuous
  Zero has an absolute meaning (Height).
  Height.
  4ft, 2ft


LABEL
TARGET,
RESPONSE
DEPENDENT
PREDICTORS
INDEPENDENT
FEATURES
y = f (X)
Nominal
ordinal
Interval
Ratio
How to
handle these
cases?
Nominal
ordinal
Int
Ratio
DISCRETE
CONTINUOUS
"CLASSIFICATION"
"REGRESSION"

PURCHASING_POWER = f(x_1 SB_AMT, x_2 AGE, x_3 GENDER, x_4 LOCATION, QUAL, JOB, FAMILY_SIZE, ... x_k)

"DISTANCE" → Imp. role in ML.

$y = f(x)$

If 'y' is known ⇒ SUPERVISED LEARNING → Regression / classification.

If 'y' is unknown ⇒ UNSUPERVISED LEARNING → Clustering.

Population → "Entire (possible) set of data points"

Sample → We can always get.

Whenever you have data, it "IS" a Sample.

What do we do with the sample?

↳ use the sample to ① understand the population
② Predict the behaviour of the pop.