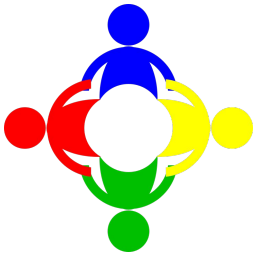
UNDERSTANDING DATA TYPES

Independent vs dependent features/variables

Features:

There are always two types of features/variables/columns:

- Independent / predictive (those that we will use to make predictions)
- Dependent / Target (**WHAT** want we want to **predict USING the INDEPENDENT FEATURES**)

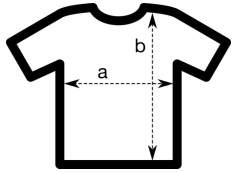


Types of features

Features:

There are always two types of features or variables:

- **Numerical**: (numbers: 1, 2, 3…,3.15, 38.4)
- **Categorical** (tags)
  - Ordinal (they display some kind of order)
  - Nominal (they don't have any order
- **Dates**:
  - Usually we use date differences.



How we can distinguish between them?

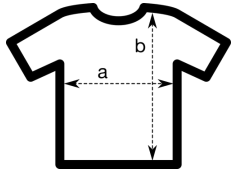
Numericals express an **amount** (they can be added, subtracted, multiplied, divided..)
Categoricals doesn't express an amount, express a tag.

Value Types found on datasets

Value types:

We can find different type of features in a dataset:

- **Numerical**:
  - Integers
  - Float numbers
- **Categorical** (tags)
  - Characters (“a”, “b”, “W”,....)
  - Strings (“My car is red”, “Yahoo!”, “englishhh”,...)
  - Booleans (True / False, Yes / Not)



Types of Machine Learning tasks

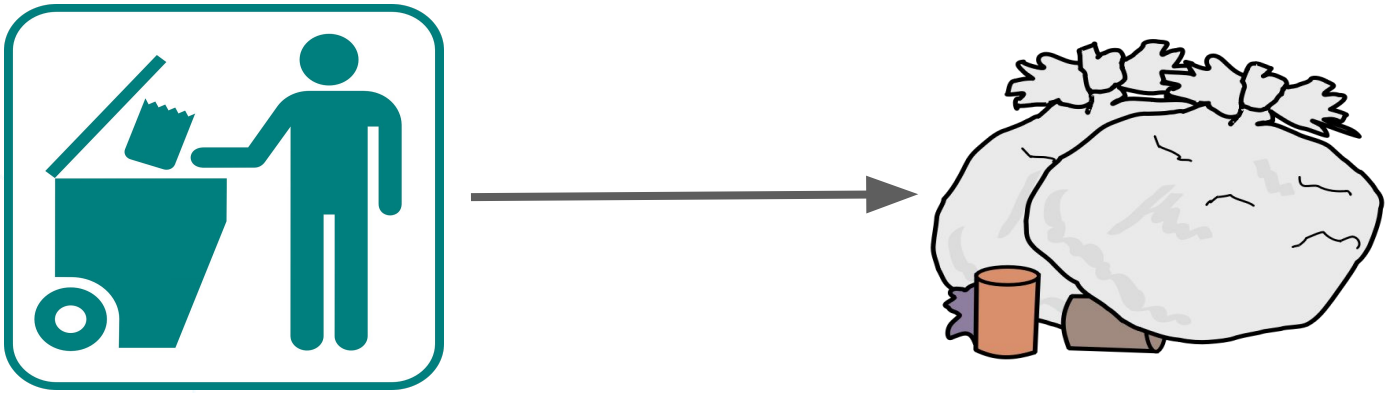
- The type of machine learning task is determined by the type of the dependent feature.
  - Numerical -> Regression task (predicting amounts)
  - Categorical -> Classification (predicting categories/labels)

- The type of task determines the what error metrics you will use to evaluate your model.

Garbage in / Garbage out

Value types:

- A model prediction will never be better than your input data!




IRON
HACK