0.1 Machine Learning

Learning is any process by which a system improves performance from experience.

Machine Learning is the study of algorithms that –

- improve their performance P
- at some task T
- with experience E.

A well-defined learning task is given by $< P, T, E >$.

0.2 Defining Learning Tasks

T : Playing checkers.

P : Percentage of games won against an arbitrary opponent.

E : Playing practice games against itself.

T : Recognizing hand-written words.

P : Percentage of words correctly classified.

E : Database of human-labeled images of handwritten words.

T : Driving on four-lane highways using vision sensors.

P : Average distance traveled before a human-judged error.

E : A sequence of images and steering commands recorded while observing a human driver.

T : Categorize email messages as spam or legitimate.

P : Percentage of email messages correctly classified.

E : Database of emails, some with human-given labels.

0.3 Why we use ML?

- Human expertise does not exist (navigating on Mars).
- Humans can't explain their expertise (speech recognition).
- Models must be customized (personalized medicine).
- Models are based on huge amounts of data (genomics).

0.4 Machine Learning Applications

- Recognizing patterns:
  - Facial identities or facial expressions.
  - Handwritten or spoken words.
  - Medical images.
- Generating pattens:
  - Generating images or motion sequences.
- Recognizing anomalies:
  - Unusual credit card transactions.
  - Unusual patterns of sensor readings in a nuclear power plant.
- Prediction:
  - Future stock prices or currency exchange rates.

0.5 Datasets and Features

- A dataset is a set of data grouped into a collection with which developers can work to meet their goals. In a dataset, rows represent the number of data points and columns represent the features of the dataset.
- The features of a dataset are the most critical aspect of the dataset, as based on the features of each available data point, will there be any possibility of deploying models to find the output to predict the features of any new data point that may be added to the dataset.

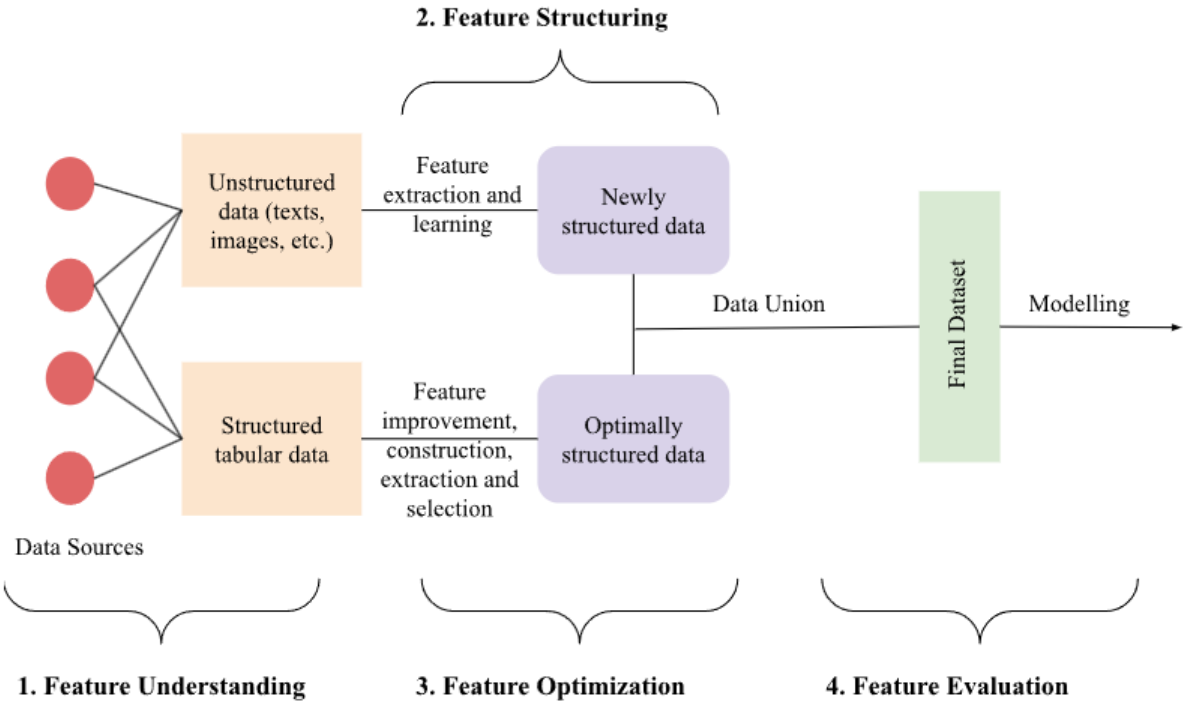

Figure 1: Datasets and Features

0.6 Feature Scaling

Scale the data to a fixed range $[0, 1]$.

- Normalization: Rescale the data x using the mean (μ) and the standard deviation (σ) of the data.

$$x_{norm} = \frac{x - \mu}{\sigma} \tag{1}$$

- Min-Max Scaling:

$$x_{minmax} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{2}$$

0.7 Types of Datasets

- Numerical Dataset.
- Categorical Dataset.
- Web Dataset.
- Time series Dataset.
- Image Dataset.

- Ordered Dataset.
- Bivariate Dataset.
- Multivariate Dataset.

0.8 The Task, T

Tasks are usually described in terms of how the machine learning should process an example: $x \in \mathbb{R}^n$ where each entry x_1 is a feature.

Classification: Learn $f : \mathbb{R}^n \to \{1, \ldots, k\}$

- $y = f(x)$: assigns input to the category with output y.
- Example: Object recognition

Regression: Learn $f : \mathbb{R}^n \to \mathbb{R}$

- Example: Weather prediction, real estate price prediction.