Advanced Institute for Artificial Intelligence

# Introduction to Machine Learning

Raphael Cóbe

raphaelmcobe@gmail.com

# Machine Learning

# Links and References

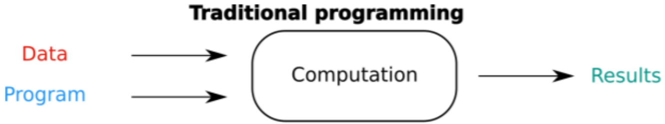
- Book: *Artificial Intelligence: A Modern Approach*
- Book: *An Introduction to Machine Learning*
- Book: *Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2*
- Machine Learning Tutorial
- Machine Learning Tutorial 2
- Video Tutorial: Supervised vs. Unsupervised Learning

## Definitions

- Science (or art) of computer programming so that they can **learn from data**;
- "*Field of study that gives computers the ability to learn without being explicitly programmed*". Arthur Samuel, 1959
- A deterministic algorithm has clear rules to return results according to the provided input.
- If the input can vary widely, this set of rules will be very large, making the execution time unfeasible.

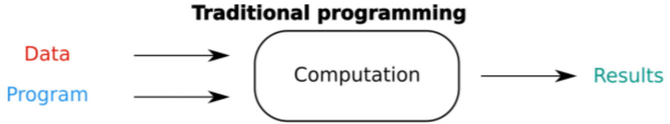
## "Traditional" Programming (Rule-Based Systems)

- Dynamic nature of problems requires constant redefinition of rules
- Email SPAM detection system
  - E.g., a machine learning-based spam filter is capable of using various criteria for such classification
    - Characterization of a SPAM can be dynamically adapted according to user markings
    - *Spammers* identify that rules do not detect numbers and change "Two" to 2



## “Traditional” Programming (Rule-Based Systems)

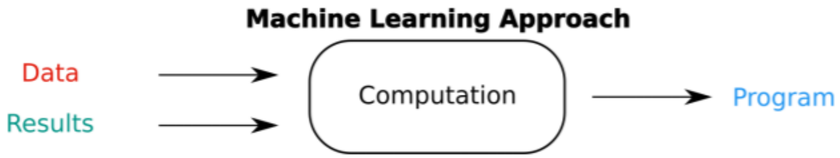
- Dynamic nature of problems requires constant redefinition of rules
- Email SPAM detection system
  - E.g., a machine learning-based spam filter is capable of using various criteria for such classification
    - Characterization of a SPAM can be dynamically adapted according to user markings
    - *Spammers* identify that rules do not detect numbers and change "Two" to 2




- **Every small change will require rule adaptation.**

# Machine Learning

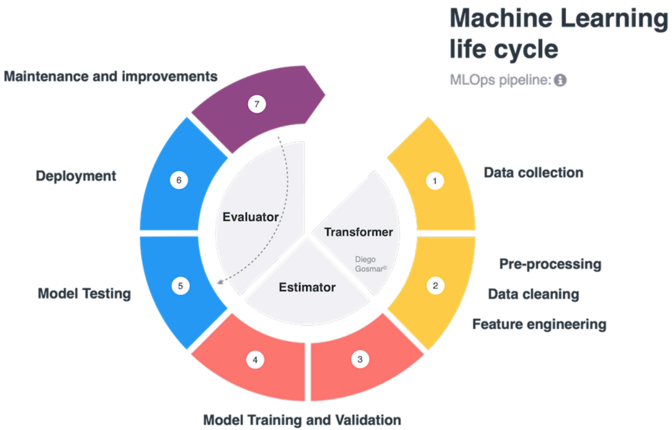
- Fundamentally involves building mathematical models to help understand data
  - Arbitrarily complex functions
- Parameter adjustments
  - Allows models to be adapted to observed data
- Thus, such models can be used to predict and understand aspects of unknown data



## Utilization of Machine Learning

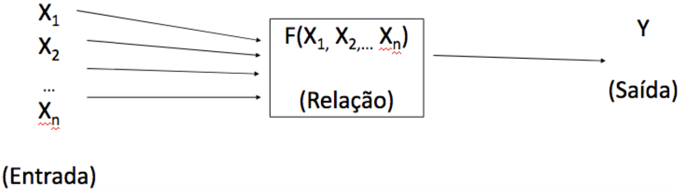
- Algorithms can be improved based on result analysis;
- Application of techniques to evaluate large amounts of data
  - Discovering patterns that were not apparent
- Used as an iterative process, seeking solutions from data, and optimizing the use of data and algorithms
- This process can be automated to some extent;

## Development Cycle



## Statistical Learning

- Until the 1990s, it was a problem of estimating a function from a given data collection;
- With the development of new analysis techniques in the 1990s (e.g., *Support Vector Machines*)
  - Not only a tool for theoretical analysis
  - Tool for creating practical algorithms to estimate **functions with inputs in N-Dimensions**;



## How to estimate the function $f$?

- The statistical process starts from a set of known events
  - Training set
- Each event has one or more predictor variable values $\mathcal{X} : X_1, X_2, ..., X_n$ and an output value $\mathcal{Y}$
- Evaluation of function $f$ performance
- Distance between the predicted value and the observed value $\varepsilon$
- Use *statistical learning* on the training set to estimate function $f$;
  - Find a function $\hat{f}$ such that $\mathcal{Y} \approx \hat{f}(\mathcal{X})$ for any observation $(\mathcal{X}, \mathcal{Y})$

## Why estimate the function $f$?

- Prediction: estimate the value of an output variable $\mathcal{Y}$ from one or more input variable values $\mathcal{X}$
  - Taking into account future data (i.e., unseen by the model - for which we do not know the value $\mathcal{Y}$)
- Inference: understand the relationship between each variable $\mathcal{X}$ and variable $\mathcal{Y}$ - how changes in $X_1, ..., X_n$ affect the value of $\mathcal{Y}$
  - *Which predictors are associated with the response?*
  - *What is the relationship between the response and each predictor?*

## Elementary Categories of Machine Learning Algorithms

- Supervised
  - Classification
  - Regression
- Unsupervised
  - Clustering
  - Dimensionality Reduction
- Semi-Supervised
  - Generative Models

# Machine Learning

## Supervised Learning

## Supervised Learning

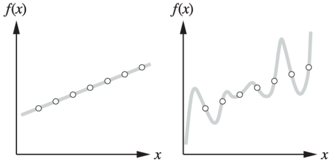
- Involves modeling the relationship between data's characteristic measures and some associated data label
- The determined model can be used to apply labels to new data
- Types of supervised algorithms
  - Classification: labels are discrete categories
  - Example of spam filter: Emails are marked as spam or non-spam. Model classifies new emails
  - Regression: labels are continuous quantities
  - Example: predicting the price of a car considering a set of predictor variables (mileage, age, brand)

## Supervised Learning (cont.)

- Given a training set with $N$ examples of input-output pairs $(\mathcal{X}_1, y_1), (\mathcal{X}_2, y_2), \ldots, (\mathcal{X}_N, y_N)$
  - Each $y_i$ is generated by an unknown function $y = f(x)$;
- The function $\hat{f}$ is called a hypothesis;
- Learning is a **search in the space of possible hypotheses** that will have **good performance**, even on new examples **beyond the training set**;
- To measure the **accuracy of a hypothesis**, we provide a set of **test examples** that are **distinct from the training set**
  - A hypothesis **generalizes well** if it predicts the $y$ value correctly for **new examples**
- $f$ can be stochastic - not strictly a function of $\mathcal{X}$
  - Learning the conditional probability distribution, $P(\mathcal{Y}|\mathcal{X})$.

## Supervised Learning (cont.)

- Hypothesis space $\mathcal{H}$
- A consistent hypothesis agrees with all the data;

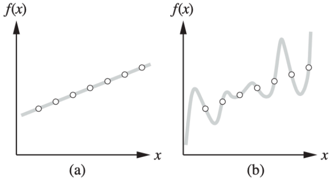



(a) (b)

- *How can we choose between various consistent hypotheses?*

## Supervised Learning (cont.)

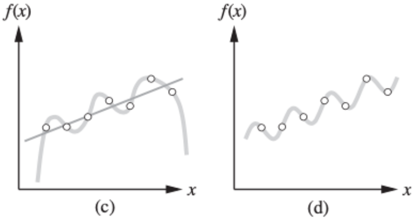
- Hypothesis space $\mathcal{H}$
- A consistent hypothesis agrees with all the data;




- *How can we choose between various consistent hypotheses?*
- **Ockham's razor**

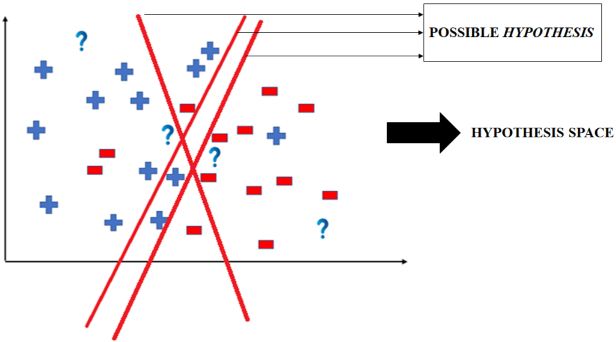
## Supervised Learning (cont.)

- Choosing the hypothesis space:
- Polynomial in $\mathcal{X}$ vs $sin(\mathcal{X})$



## Supervised Learning (cont.)

- In the case of classification:



## Classification vs Regression

- *In a nutshell*:
  - Classification is the task of predicting a discrete class label.
  - Regression is the task of predicting a continuous quantity.
- There's some overlap between classification and regression algorithms; for example:
  - A classification algorithm can predict a continuous value, but the continuous value is in the form of a probability for a class label.
  - A regression algorithm can predict a discrete value, but the discrete value in the form of an integer quantity.

## Classification vs Regression (cont.)

- Some algorithms can be used for both with slight modifications
  - Decision trees and artificial neural networks;
- How we evaluate classification and regression predictions vary and **do not overlap**
  - Classification predictions can be evaluated using **accuracy**, while regression predictions cannot.
  - Regression predictions can be evaluated using **root mean squared error** (RMSE), while classification predictions cannot.

## Key Characteristics

- For any problem to be investigated as Machine Learning, we have some common characteristics:
  - Samples: rows in the dataset
  - Features: columns in the dataset
  - Feature Matrix: Combination of rows and features
  - Target vector: column to be predicted

## Key Characteristics (cont.)

- Machine Learning algorithms usually require a large amount of data to provide a satisfactory solution
- Data needs to be representative concerning the problem being investigated
- Consider the influence of categories in relation to the complete dataset
- Data Quality:
  - Consider detecting and, if possible, eliminating outliers and noise
  - Discard redundant data
  - They are unnecessary when placed in the context of another attribute
  - E.g., Social class and monthly income
  - Discard irrelevant data
  - They have no relation to the target attribute
  - E.g., Social Security Number and disease

## Iterative Machine Learning Design

- Define the problem to be tackled with a predictive model
- Organize data according to the defined problem
- Define an evaluation metric
- Split the data into training and testing according to the metric
- Inspect the solution
- Propose improvements to the model or data organization

- The process of organizing data according to the defined model involves the following activities:
  - Exchange categorical or ordinal data for numbers
  - Change the scale of the data
  - Eliminate missing values or replace them with another value
  - Separate predictor variables and target variables
  - Split the dataset into training and testing