01-Introduction

Why machine learning?

- Computer/Computing science aims to develop **automated** machinery (i.e., programs) to accomplish non-trivial tasks. This is known as **programming**.
- A large number of tasks cannot be programmed explicitly by humans
  - E.g., spam detection, hand written digit recognition
  - Such programs **do exist**, but humans cannot program explicitly
  - Some machinery (e.g., a machine learning model) is able to yield a program that almost accomplishes the task.

What is machine learning?

$\text{Data} \xrightarrow{\text{Machine learning Algorithm}} \text{Model}$

$\text{Possibly new Data} \longrightarrow \text{Prediction}$

Note: Prediction is more important than a machine learning model

C.f.: Human learning

E.g.: Predict the stock price of tomorrow

- A human without any prior knowledge of stocks is unable to predict the stock price
- Some kind of human learning is needed

Data = history $\xrightarrow{\text{Human learning}}$ Model =

New Data = Snapshot of Today's world $\longrightarrow$ Prediction = Tomorrow's stock price

Machine learning example – Supervised learning (Regression)

Training/Learning:

Some representation of history

e.g.,

$\text{Data} = \boxed{\{(x_1^{(i)}, x_2^{(i)}), y^{(i)}\}_{i=1}^{M}} \xrightarrow{\text{Machine learning}} \text{Model} =$

Inference/ Prediction:

New Data = $\boxed{(x_1^*, x_2^*)}$ $\longrightarrow$ Prediction = Tomorrow's stock price

$(x_1, x_2)^T \in \mathbb{R}^2$, $y \in \mathbb{R}$ $\qquad y^* = ?$

Representations

Features (the main approach of this course)
- The stock price of today
- The stock price of yesterday
- The standard deviation of today
- Etc.

Similarity of data samples
- E.g., prediction by nearest neighbor

Raw data in certain applications
- E.g., pixels in image processing, words in natural language processing

Measure of Success

- Stock prediction: The gain/loss of money |\$actual - \$predicted|
- Sentiment classification: fraction of correct predictions (accuracy)
- Cancer detection

Machine learning example – Supervised learning (Classification)

Some representation of history

e.g.,

Data = $\boxed{\{(x_1^{(i)}, x_2^{(i)}), y^{(i)}\}_{i=1}^{M}}$ $\xrightarrow{\text{Machine learning}}$ Model =

New Data = $\boxed{(x_1^*, x_2^*)}$ $\longrightarrow$ Prediction = Tomorrow's stock price

$y^* = ?$

Feature $x_1 \in [0, 1]$: % of out of vocabulary words

Feature $x_2 \in \{0, 1\}$: If the word "money" occurs

$y \in \{0, 1\}$: "1" means not spam, "0" means spam

Heuristics vs Machine learning

- Heuristics/rule-based

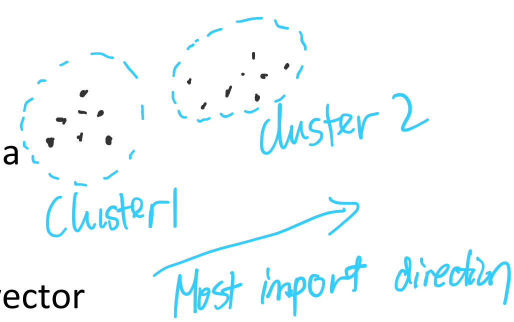
- Heuristics, rule based
    - Humans define a rule: if x_1>=0.2 and x_2=1, then y=1
    - Machine uses the rule(s) to predict spam
      ==> NOT machine learning
- Machine learning:
    - Humans specify a set of "models"
      E.g., $\text{score} := w_1 \cdot x_1 + w_2 \cdot x_2$, $y = \begin{cases} 1 & \text{if score} \geq b \\ 0 & \text{if score} < b \end{cases}$
      $(w_1, w_2, b)^T \in \mathbb{R}^3$
    - Machine observes data $\{(x_1^{(i)}, x_2^{(i)}), y^{(i)}\}_{i=1}^{M}$
    - Machine learns how features x_1, x_2 play a role in the task
      E.g., w_1=0.2, w_2=0.9, b=1
      Note: If humans define the weights, then NOT machine learning
    - Machine predicts spam based on the learned model

Regression vs Classification

- Regression: y is a real number
    - Implicit distance and order among different values of y
- Classification: y is a discrete categorical variable
    - No order is defined. Distance is trivial
- Some tasks can be modeled as either classification or regression
    - Sentiment prediction among 1 ... 5
    - Classification: Treat 1, 2, ..., 5 as individual categories
    - Regression
        - Predict a real-valued sentiment score
        - Ground the score to the nearest integer

Machine learning examples - Unsupervised learning

- Clustering
- Outlier detection
- Dimensionality reduction
    - Representing high-dimensional data in a low dimensional space
- Vector representation learning
    - Representing an input data point as a vector
    - Useful for other (downstream) tasks



Machine learning examples - Reinforcement learning

- Robotics
- Alpha Go

- Alpha Go

Other classification criteria

- Batch learning vs. Online learning
  - Batch: All data available before training
  - Online: Data samples come one after another
- Passive learning vs. Active learning
  - Passive learning: Only observing given data
  - Active learning: Can ask labels for some data as the learning machine wants
- Structured prediction
  - y itself has internal structures, e.g., a sequence, a tree.

This course mainly focuses on supervised, batch, passive, non-structured

<u>Tentative syllabus</u>

Linear regression
- Mean square error (as heuristics)
- Closed-form solution
- Gradient descent
- Maximum likelihood estimation
- Maximum a posteriori training
- Bias-variance tradeoff
- Train-validation-test framework
- Bayesian learning
- Generalized linear models

Linear classification
- Discriminative model: Logistic regression
- Multi-class softmax
- Maximum a posteriori inference
- Generative model: Naïve Bayes
- Discriminant: Linear SVM

Nonlinear models
- Kernels methods
- Neural networks