Artificial Intelligence (CS303)

Lecture 4: Principles of Machine Learning

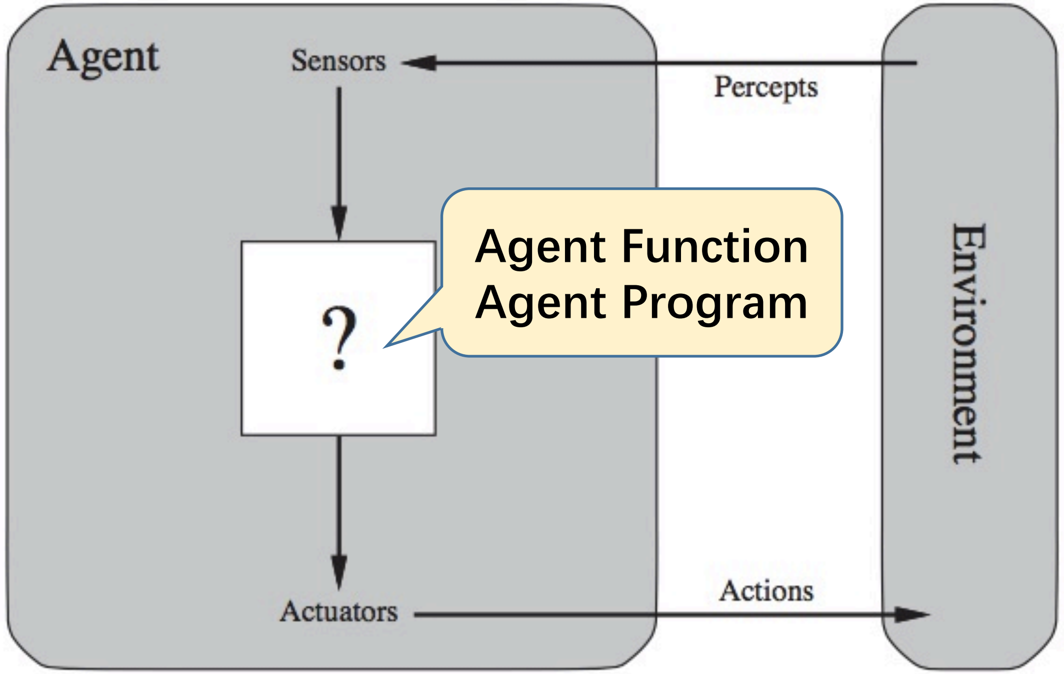
Hints for this lecture

- **Learning = Search for hypothesis/functions that generalize well.**

Outline of this lecture

- **What is Learning**
- **Key Questions for Learning**
- **Learning Paradigms and Principles**

Why Learning is Important for AI?



- Who designed/implemented the search algorithm behind AlphaGo?

Why Learning is Important for AI?

- Intuitively, human behavior is not a rigid/static program, i.e., we might behave different as growing up.

- Scientifically, Learning is universal and a major source of our behavioral change.

- From an engineering perspective, it is also intractable to implement an agent function/program that encompasses all possible input/output pairs in the complex world. Thus an AI with learning ability would also help.

What is Learning (in AI)?

- Examples for human learning
  - Learn similarity between objects
  - Learn to recognize objects
  - Learn to pass the college entrance examination

- **Machine Learning**: Given some observations (data) from the environment, how could an agent improve its agent function?

- Intuitive assumptions
  - the data share something in common
  - “something” could be obtained by an algorithm/program

To Learn What?

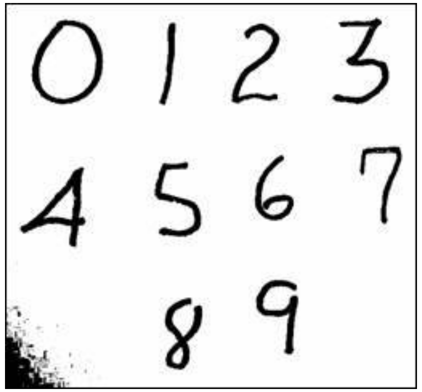
- Ideally, the purpose of (Machine) Learning is to achieve a “universal” agent function that can give the appropriate output for any input that we can imaging.

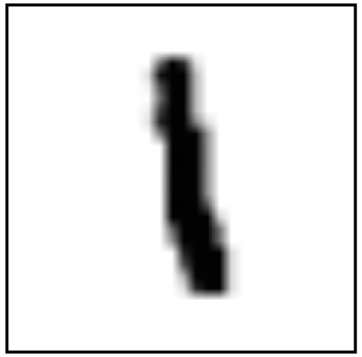
- Unfortunately, getting a universal function is impractical (at least for now).

- To be realistic, we can focus on simpler learning tasks, e.g., the ability of recognizing a hand-written digit should be much easier to learn than the ability to make a good living.

An example

- Hand-written Digit Recognition
  - Data: image of a hand-written digit
  - Agent function: a function mapping an image to a digit
  - Improvement: how many images can be recognized correctly.





$$\approx \begin{bmatrix}
0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&.6&.8&0&0&0&0&0&0\\
0&0&0&0&0&0&.7&1&0&0&0&0&0&0\\
0&0&0&0&0&0&.7&1&0&0&0&0&0&0\\
0&0&0&0&0&0&.5&1&.4&0&0&0&0&0\\
0&0&0&0&0&0&0&1&.4&0&0&0&0&0\\
0&0&0&0&0&0&0&1&.4&0&0&0&0&0\\
0&0&0&0&0&0&0&1&.7&0&0&0&0&0\\
0&0&0&0&0&0&0&1&1&0&0&0&0&0\\
0&0&0&0&0&0&0&.9&1&.1&0&0&0&0\\
0&0&0&0&0&0&0&.3&1&.1&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0
\end{bmatrix}$$

transform to a vector

A Naive Parametric Method

- Classify a data to the class with the highest posteriori probability

$$P(w_j|x) = \frac{p(x|w_j)P(w_j)}{p(x)}$$

$$P(w_2|x) > P(w_1|x) \quad \Longleftrightarrow \quad \ln p(x|w_i) + \ln P(w_i) > \ln p(x|w_i) + \ln P(w_i)$$

- Assumption: data follows independent identically distribution

A Naive Parametric Method

- In addition to the i. i. d assumption, further assume the data follows a specific distribution (e.g., Gaussian distribution)

- Estimate the prior and the likelihood.

- “parametric”: the assumption on the probability density function.

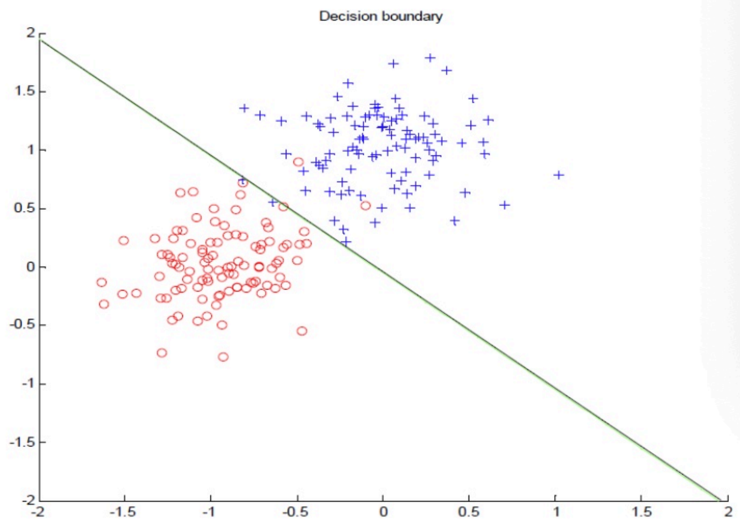
- Parametric methods usually do not involve parameters to fine-tune, while Nonparametric methods usually do.

Agent as A Linear Function

- Find a straight line/hyper-plane to separate datum from different classes.

$$g(\boldsymbol{x}) = \boldsymbol{w}^t \boldsymbol{x} + b$$

input space $\mathcal{X} \subset \mathbb{R}^d$ and output space $\mathcal{Y} \in \{0, 1\}$



Key Questions for Machine Learning

- Given some observations (data) from the environment, how could an agent improve its agent function?
  - What is the format of the data? (data representation)
  - What does the agent function look like? (model representation)
  - How to measure the "improvement"? (objective function)
  - What is the learning algorithm? (to get a good agent function)

Representation + Algorithm + Evaluation = Agent function/Model

Search in a model/hypothesis space

Learning Principles

Generalization, Generalization, Generalization: the learned agent function is expected to be able to handle previously unseen situations.（举一反三，不要刻舟求剑）

- How to calculate/estimate the generalization?
  - Design an appropriate objective function for learning
- Overfitting: the model perfectly fit the seen data perfectly, but cannot generalize well.
  - Occam’s Razor: An explanation of data should be made as simple as possible, but no simpler, i.e., model complexity should be controlled.

Learning Paradigms

A Machine Learning process typically involves two phases

- Training: build the agent function
- Testing/Inference: test the agent function/deploy the agent function in real use.

Different ML techniques may use different training/learning paradigms

- Supervised Learning: the correct answer is available to the learning algorithm.
- Reinforcement Learning: the only feedback is the reward of an output, e.g., the output is correct or not (the correct answer is not given).
- Unsupervised Learning: no correct answer is available.

Summary

- Machine learning is basically the search for a model/hypothesis/rule that can map an input from the environment to an appropriate output, such that we don't need to program them in advance.

- Different settings of learning may lead to different learning algorithms.

- Generalization is the ultimate goal of learning, but usually uneasy to measure.

- In this course, we mainly use classification problem as an example:
  - Classification is a sufficiently abstract and have numerous applications.
  - It has so far led to many successful applications of machine learning and AI.

What is classification?

- Given a set of class labels and a set of training data, usually represented by a set of features, to achieve a classifier that (ideally) can assign the correct label to any previous unseen data.

ID	Height	Weight
John	1.75米	80KG
Mike	1.8米	75KG

- In most literature, classification refers to supervised learning, i.e., labels for the training data are given to the learning algorithm.

To be continued