Machine Learning - 4105931

Lecture 2
The Learning Problem

Chen-Kuo Chiang (江 振 國)
ckchiang@cs.ccu.edu.tw
中正大學 資訊工程學系

課程教材

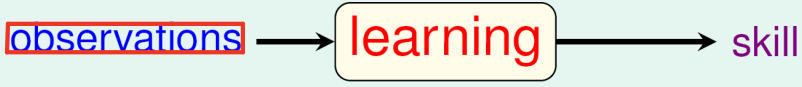
- Machine Learning Foundations
  - http://www.csie.ntu.edu.tw/~htlin/mooc/
- Text Book
  - Learning from Data, Yaser Abu-Mostafa, Malik Magdon-Ismail and Hsuan-Tien Lin.
- Online Course Materials (機器學習基石)
  - https://www.youtube.com/playlist?list=PLXVfgk9fNX2I7tB6oIINGBmW50rrmFTqf

Section Summary

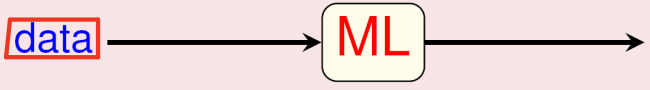
- When Can Machines Learn?
  - Lecture 1: The Learning Problem
    - What is Machine Learning
    - Applications of Machine Learning
    - Components of Machine Learning
    - Machine Learning and Other Fields

From Learning to Machine Learning

learning: acquiring skill
with experience accumulated from observations



machine learning: acquiring skill
with experience accumulated/**computed** from data

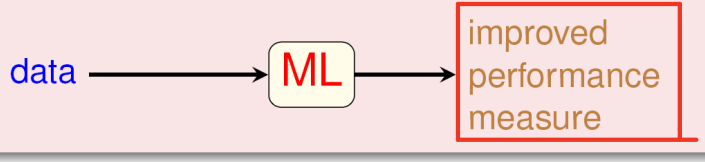


What is skill?

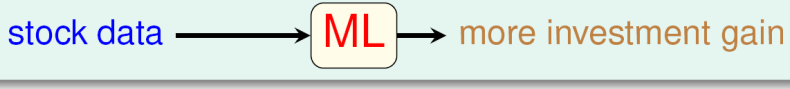
A More Concrete Definition

skill
$\Leftrightarrow$ improve some performance measure (e.g. prediction accuracy)

machine learning: improving some performance measure
with experience **computed** from data



An Application in Computational Finance

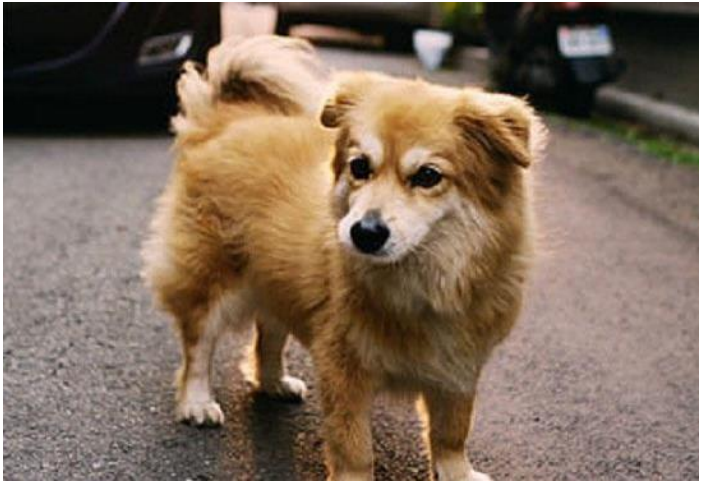


Why use machine learning?

Dog Recognition

- 如何定義柴犬? (Instagram : shiba_feifei) 柴犬飛飛

Are These Shiba Dogs?



Tree Recognition

- ‘Define’ Shiba Dogs : **difficult**.
- Learn from data (observations) and recognize: a **3-year-old can do so**.
- ‘ML-based Shiba Dog recognition system’ can be **easier to build** than hand-programmed system.

ML: an **alternative route** to build complicated systems

Use Cases of Using Machine Learning

ML: an **alternative route** to build complicated systems

Some Use Scenarios

- when human cannot program the system manually
  —navigating on Mars 系統的複雜
- when human cannot 'define the solution' easily
  —speech/visual recognition 答案不容易定義
- when needing rapid decisions that humans cannot do
  —high-frequency trading 快速決策
- when needing to be user-oriented in a massive scale
  —consumer-targeted marketing 使用者

Key Essence of Machine Learning

- 使用機器學習的時機

1. exists some 'underlying pattern' to be learned
   —so 'performance measure' can be improved 資料分佈需要規則
2. but no programmable (easy) definition
   —so 'ML' is needed 複雜度高
3. somehow there is data about the pattern
   —so ML has some 'inputs' to learn from 收集得到資料

key essence: help decide whether to use ML

Fun Time

Which of the following is best suited for machine learning?

1. predicting whether the next cry of the baby girl happens at an even-numbered minute or not 隨機/沒有資料分佈
2. determining whether a given graph contains a cycle
3. deciding whether to approve credit card to some customer
4. guessing whether the earth will be destroyed by the misuse of nuclear power in the next ten years 沒有資料

Fun Time

Which of the following is best suited for machine learning?

1. predicting whether the next cry of the baby girl happens at an even-numbered minute or not
2. determining whether a given graph contains a cycle
3. deciding whether to approve credit card to some customer
4. guessing whether the earth will be destroyed by the misuse of nuclear power in the next ten years

Reference Answer: (3)

1. no pattern
2. programmable definition
3. pattern: customer behavior;
   definition: not easily programmable;
   data: history of bank operation
4. arguably no (or not enough) data yet

Applications of Machine Learning

- Daily Needs: Food, Clothing, Housing, Transportation

1. Food (Sadilek et al., 2013)
   - data: Twitter data (words + location)
   - skill: tell food poisoning likeliness of restaurant properly
2. Clothing (Abu-Mostafa, 2012)
   - data: sales figures + client surveys
   - skill: give good fashion recommendations to clients
3. Housing (Tsanas and Xifara, 2012)
   - data: characteristics of buildings and their energy load
   - skill: predict energy load of other buildings closely
4. Transportation (Stallkamp et al., 2012)
   - data: some traffic sign images and meanings
   - skill: recognize traffic signs accurately

ML is everywhere!

Components of Learning

- Credit Approval

Applicant Information

age	23 years
gender	female
annual salary	NTD 1,000,000
year in residence	1 year
year in job	0.5 year
current debt	200,000

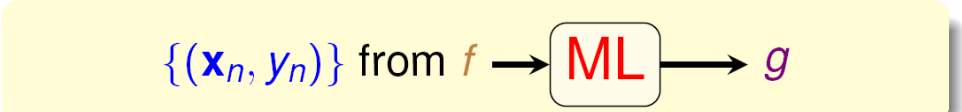
unknown pattern to be learned:

‘approve credit card good for bank?’

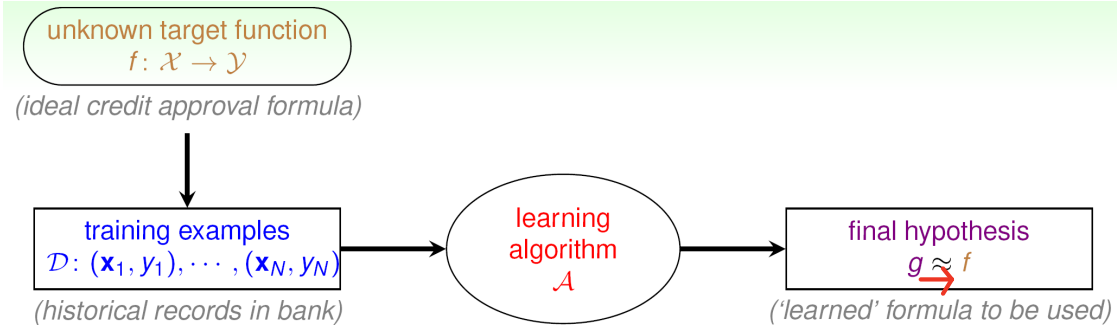
Formalize the Learning Problem

Basic Notations

- input: $\mathbf{x} \in \mathcal{X}$ (customer application)
- output: $y \in \mathcal{Y}$ (good/bad after approving credit card)
- unknown pattern to be learned $\Leftrightarrow$ target function:
  $f : \mathcal{X} \to \mathcal{Y}$ (ideal credit approval formula)
- data $\Leftrightarrow$ training examples: $\mathcal{D} = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \cdots, (\mathbf{x}_N, y_N)\}$
  (historical records in bank) 找出分類模型
- hypothesis $\Leftrightarrow$ skill with hopefully good performance:
  $g : \mathcal{X} \to \mathcal{Y}$ ('learned' formula to be used)



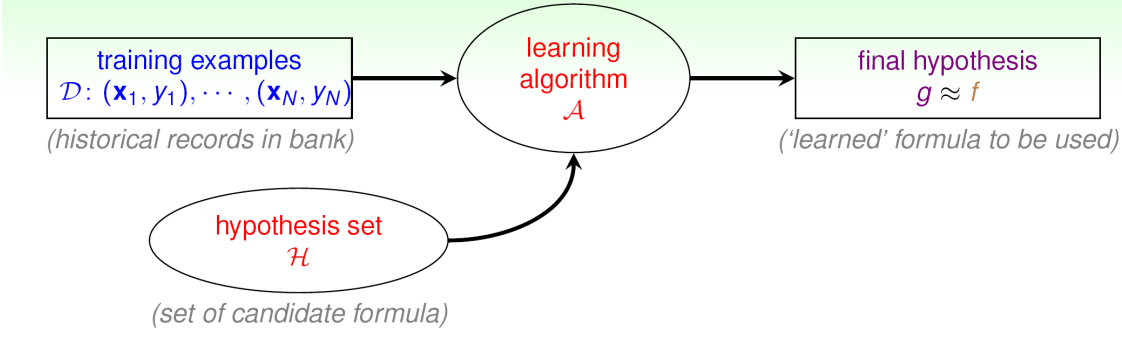
Learning Flow



- target f **unknown** (i.e. no programmable definition)
- hypothesis g hopefully $\approx f$ but possibly **different** from f (perfection 'impossible' when f unknown)

What does g look like?

The Learning Model

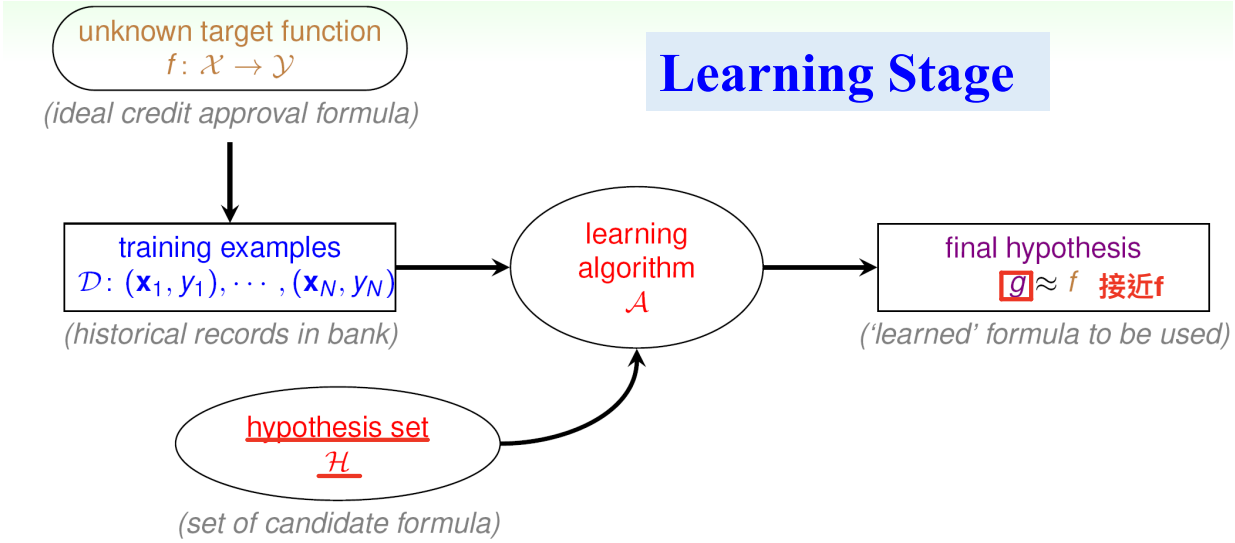


- assume $g \in \mathcal{H} = \{h_k\}$, i.e. approving if
  - h_1: annual salary > NTD 800,000
  - h_2: debt > NTD 100,000 (really?)
  - h_3: year in job $\leq$ 2 (really?)
- hypothesis set $\mathcal{H}$:
  - can contain **good or bad hypotheses**
  - up to $\mathcal{A}$ to pick the 'best' one as g

找出g最接近f

learning model = $\mathcal{A}$ and $\mathcal{H}$

Practical Definition of Machine Learning



machine learning:
use data to compute hypothesis g
that approximates target f

Practical Definition of Machine Learning

Testing Stage

machine learning: acquiring skill
with experience accumulated/**computed** from data

data → ML → skill

Fun Time

How to use the four sets below to form a learning problem for song recommendation?

$\mathcal{S}_1$ = $[0, 100]$

$\mathcal{S}_2$ = all possible (userid, songid) pairs 所有可能組合

$\mathcal{S}_3$ = all formula that 'multiplies' user factors & song factors, indexed by all possible combinations of such factors 所有可能公式

$\mathcal{S}_4$ = 1,000,000 pairs of ((userid, songid), rating) 評分

1. $\mathcal{S}_1 = \mathcal{X}, \mathcal{S}_2 = \mathcal{Y}, \mathcal{S}_3 = \mathcal{H}, \mathcal{S}_4 = \mathcal{D}$
2. $\mathcal{S}_1 = \mathcal{Y}, \mathcal{S}_2 = \mathcal{X}, \mathcal{S}_3 = \mathcal{H}, \mathcal{S}_4 = \mathcal{D}$
3. $\mathcal{S}_1 = \mathcal{D}, \mathcal{S}_2 = \mathcal{H}, \mathcal{S}_3 = \mathcal{Y}, \mathcal{S}_4 = \mathcal{X}$
4. $\mathcal{S}_1 = \mathcal{X}, \mathcal{S}_2 = \mathcal{D}, \mathcal{S}_3 = \mathcal{Y}, \mathcal{S}_4 = \mathcal{H}$

Reference Answer: (2)

$$\mathcal{S}_4 \xrightarrow{\mathcal{A} \text{ on } \mathcal{S}_3} (g : \mathcal{S}_2 \to \mathcal{S}_1)$$

Machine Learning and Data Mining

Machine Learning

use data to compute hypothesis g that approximates target f

Data Mining

找出有趣特性

use **(huge)** data to **find property** that is interesting

- if 'interesting property' **same as** 'hypothesis that approximate target'
  —**ML = DM** (usually what KDDCup does)
  假設是有相關
- if 'interesting property' **related to** 'hypothesis that approximate target'
  —**DM can help ML, and vice versa** (often, but not always)
- traditional DM also focuses on **efficient computation in large database**

difficult to distinguish ML and DM in reality

Data Mining 實例

- 英國倫敦基金公司Derwent Capital Markets於2011年曾利用Twitter上發表的推文去統計大眾情緒以預測股市走勢，因此在當年全球市場低迷之時，還能維持1.85%的報酬率，和S&P500下跌了2.2%的指數相比，領先許多。
- 美國零售商Target利用公司內部所擁有的消費者購買資料進行分析，並由此去預測消費者的行為—例如預測孕婦在懷孕初、中期大概會想購買甚麼樣的物品；然後當消費者在網路上購買了某一項產品，系統就會自動提供更多其可能會感興趣的產品資訊。
  - 當時Target寄送了孕婦用品廣告到有可能購買的消費者家中，其中一位收到廣告的孩子父親非常生氣，特地跑到Target去理論，認為自己女兒不需要這種產品，為何賣場要寄這種「有辱名節」的嫌疑廣告；結果事後才發現女兒是真的懷孕了。這個有意思的例子就是「可以比爸爸更早知道女兒懷孕」的資料探勘。

Machine Learning and Artificial Intelligence

Machine Learning

use data to compute hypothesis g that approximates target f

Artificial Intelligence

compute **something that shows intelligent behavior**

- $g \approx f$ is something that shows intelligent behavior
  —**ML can realize AI**, among other routes
- e.g. chess playing
  - traditional AI: game tree
  - ML for AI: 'learning from board data'

ML is one possible route to realize AI

Machine Learning and Statistics

Machine Learning

use data to compute hypothesis g that approximates target f

Statistics

use data to **make inference** 推論 **about an unknown process**

- g is an inference outcome; f is something unknown —statistics **can be used to achieve ML**
- traditional statistics also focus on **provable results with math assumptions** 較未考量計算量, and care less about computation

statistics: many useful tools for ML

Summary

Lecture 1: The Learning Problem

- What is Machine Learning
  use data to approximate target
- Applications of Machine Learning
  almost everywhere
- Components of Machine Learning
  $\mathcal{A}$ **takes** $\mathcal{D}$ **and** $\mathcal{H}$ **to get** g
- Machine Learning and Other Fields
  related to DM, AI and Stats