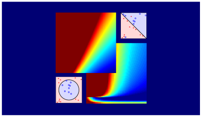
Basics of Machine Learning
(機器學習入門)



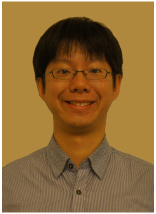
Hsuan-Tien Lin (林軒田)

htlin@csie.ntu.edu.tw

Department of Computer Science
& Information Engineering

National Taiwan University
(國立台灣大學資訊工程系)

More about Me



Associate Professor

Dept. CSIE

National Taiwan University

- graduated from ck49th314, 1997
- co-leader of KDDCup world champion teams at NTU: 2010–2013
- Secretary General, Taiwanese Association for Artificial Intelligence
- co-author of bestseller ML textbook "Learning from Data"
- instructor of Mandarin-teaching MOOC of Machine Learning on NTU-Coursera: **2013.11–**

  `https://www.coursera.org/course/ntumlone`

From Learning to Machine Learning

learning: acquiring skill
with experience accumulated from observations

$\text{observations} \longrightarrow \boxed{\text{learning}} \longrightarrow \text{skill}$

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

$\text{data} \longrightarrow \boxed{\text{ML}} \longrightarrow \text{skill}$

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

data $\longrightarrow$ $\boxed{\text{ML}}$ $\longrightarrow$ improved performance measure

An Application in Computational Finance

stock data $\longrightarrow$ $\boxed{\text{ML}}$ $\longrightarrow$ more investment gain

Why use machine learning?

Yet Another Application: Tree Recognition

- 'define' trees and hand-program: **difficult**
- learn from data (observations) and recognize: a **3-year-old can do so**
- 'ML-based tree recognition system' can be **easier to build** than hand-programmed system

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

The Machine Learning Route

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

Some Use Scenarios

- when human cannot program the system manually
  —navigating on Mars
- when human cannot 'define the solution' easily
  —speech/visual recognition
- when needing rapid decisions that humans cannot do
  —high-frequency trading
- when needing to be user-oriented in a massive scale
  —consumer-targeted marketing

Give a **computer** a fish, you feed it for a day;
teach it how to fish, you feed it for a lifetime. **:-)**

Key Essence of Machine Learning

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

data $\longrightarrow$ ML $\longrightarrow$ improved performance measure

1. exists some 'underlying pattern' to be learned
   —so 'performance measure' can be improved
2. but no programmable (easy) definition
   —so 'ML' is needed
3. somehow there is data about the pattern
   —so ML has some 'inputs' to learn from

key essence: help decide whether to use ML

Daily Needs: Food, Clothing, Housing, Transportation

data → ML → skill

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!

Education

$$\text{data} \longrightarrow \boxed{\text{ML}} \longrightarrow \text{skill}$$

- data: students' records on quizzes on a Math tutoring system
- skill: predict whether a student can give a correct answer to another quiz question

A Possible ML Solution

answer correctly $\approx$ $[\![$recent strength of student $>$ difficulty of question$]\!]$

- give ML 9 million records from 3000 students
- ML determines (**reverse-engineers**) strength and difficulty automatically

key part of the **world-champion** system from National Taiwan Univ. in KDDCup 2010

Entertainment: Recommender System (1/2)

data $\longrightarrow$ ML $\longrightarrow$ skill

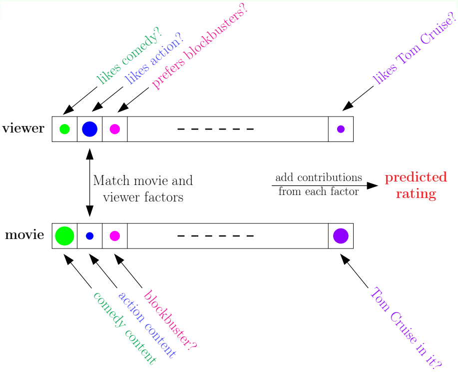
- data: how many users have rated some movies
- skill: predict how a user would rate an unrated movie

A Hot Problem

- competition held by Netflix in 2006
  - 100,480,507 ratings that 480,189 users gave to 17,770 movies
  - 10% improvement = **1 million dollar prize**
- similar competition (movies $\rightarrow$ songs) held by Yahoo! in KDDCup 2011
  - 252,800,275 ratings that 1,000,990 users gave to 624,961 songs

How can machines **learn our preferences**?

Entertainment: Recommender System (2/2)



A Possible ML Solution

- pattern:
  rating ← viewer/movie factors
- learning:
  known rating
  → learned factors
  → unknown rating prediction

key part of the **world-champion** (again!)
system from National Taiwan Univ.
in KDDCup 2011

Components of Learning: Metaphor Using 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?’

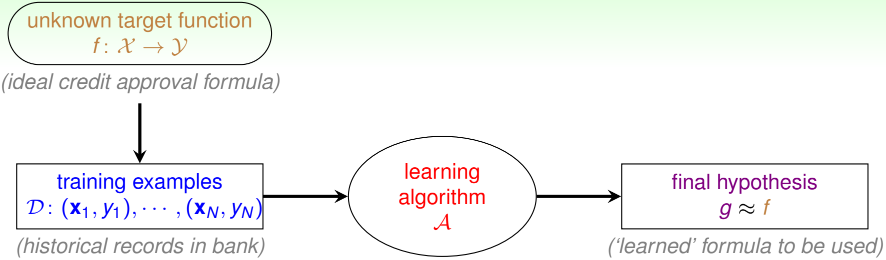
Formalize the Learning Problem

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: $f\colon \mathcal{X} \to \mathcal{Y}$ (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)
- hypothesis $\Leftrightarrow$ skill with hopefully good performance: $g\colon \mathcal{X} \to \mathcal{Y}$ ('learned' formula to be used)

$\{(\mathbf{x}_n, y_n)\}$ from f $\longrightarrow$ $\boxed{\text{ML}}$ $\longrightarrow$ g

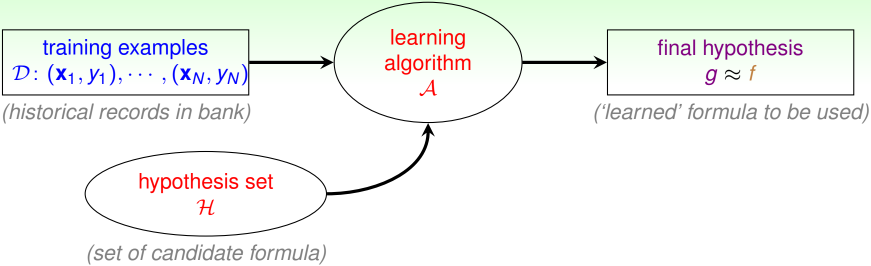
Learning Flow for Credit Approval



- 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?

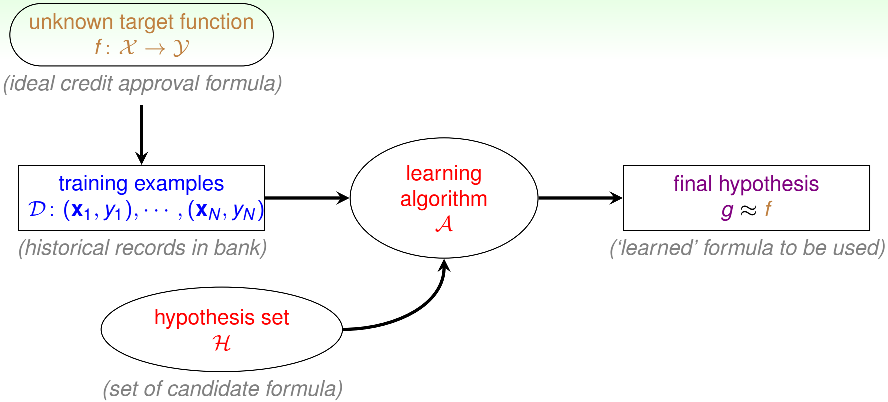
The Learning Model



- 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 ≤ 2 (really?)
- hypothesis set $\mathcal{H}$:
  - can contain **good or bad hypotheses**
  - up to $\mathcal{A}$ to pick the 'best' one as g

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

Practical Definition of Machine Learning



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

Machine Learning and Data Mining

Machine Learning

use data to compute hypothesis g that approximates target f

Data Mining

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

- if 'interesting property' **same as** 'hypothesis that approximate target'
  —**ML = DM** (usually what KDDCup does)
- 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

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

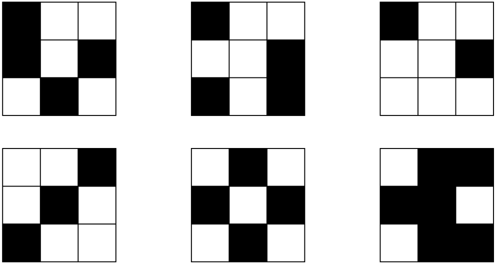
Statistics

use data to **make inference 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**, and care less about computation

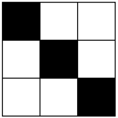
statistics: many useful tools for ML

A Learning Puzzle



$y_n = -1$

$y_n = +1$

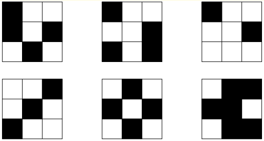


$g(\mathbf{x}) = ?$

let's test your 'human learning' with 6 examples :-)

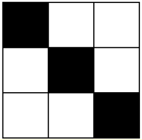
Two Controversial Answers

whatever you say about $g(\mathbf{x})$,



$y_n = -1$

$y_n = +1$



$g(\mathbf{x}) = ?$

truth $f(\mathbf{x}) = +1$ because . . .

- symmetry $\Leftrightarrow$ +1
- (black or white count = 3) or (black count = 4 and middle-top black) $\Leftrightarrow$ +1

truth $f(\mathbf{x}) = -1$ because . . .

- left-top black $\Leftrightarrow$ -1
- middle column contains at most 1 black and right-top white $\Leftrightarrow$ -1

all valid reasons, your **adversarial teacher** can always call you '**didn't learn**'. **:-(**

No Free Lunch Theorem

Without any assumptions on the learning problem on hand, all learning algorithms perform the same.

No algorithm is better for all learning problems

Gender Classification Problem

?

Male

Female

Gender Classification: Lesson 1

?

Male

Female

Female

Male

Male

Female

Female

Male

Female

Male

Gender Classification: Lesson 2

Male

Male

Female

Female

Male

Male

Female

Female

Male

Female

Male

Gender Classification: Lesson 3

Male

Male

Male

Female

Female

Male

Male

Female

Female

Male

Female

Male

Gender Classification: Lesson 4

?

Male

Female

Female

Male

Male

Female

Female

Male

Female

Male

Nearest Neighbors

Intuition

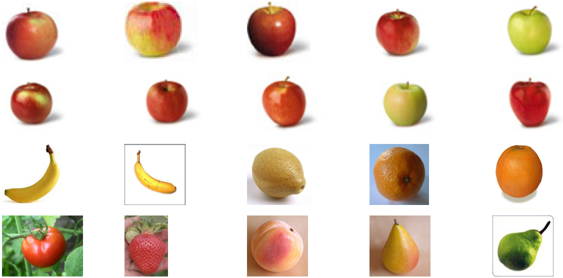
- memorize everything
- predict with the closest case

Algorithm

- Training: memorize all examples (picture, label)
- Prediction:
  - find K nearest neighbors
  - let them vote!

Apple Recognition Problem

- Is this a picture of an apple?
- We want to teach a class of 6 year olds.
- Gather photos from NY Apple Asso. and Google Image.

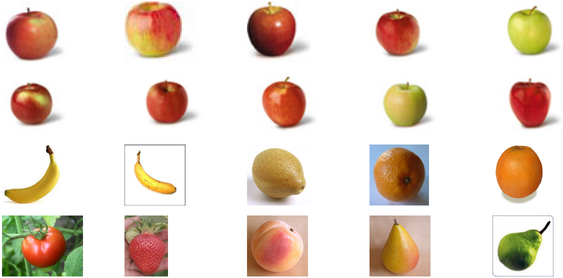


Our Fruit Class Begins

Teacher: How would you describe an apple? Michael?

Michael: I think apples are circular.

(Class): Apples are circular.

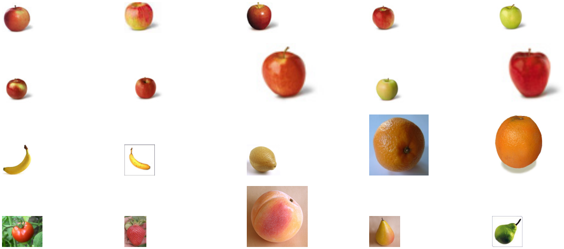


Our Fruit Class Continues

Teacher: Being circular is a good feature for the apples. However, if you only say circular, you could make several mistakes. What else can we say for an apple? Tina?

Tina: It looks like apples are red.

(Class): Apples are somewhat circular and somewhat red.

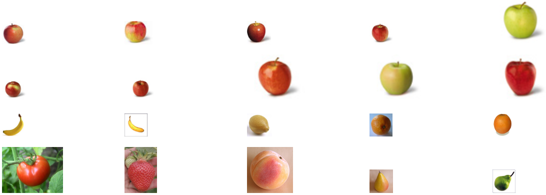


Our Fruit Class Continues

Teacher: Yes. Many apples are red. However, you could still make mistakes based on circular and red. Do you have any other suggestions, Joey?

Joey: Apples could also be green.

(Class): Apples are somewhat circular and somewhat red and possibly green.

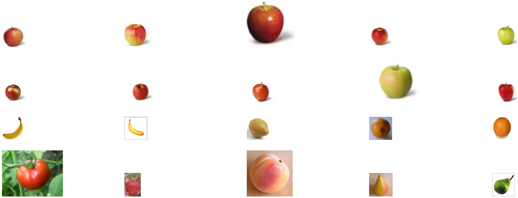


Our Fruit Class Continues

Teacher: Yes. It seems that apples might be circular, red, green. But you may confuse them with tomatoes or peaches, right? Any more suggestions, Jessica?

Jessica: Apples have stems at the top.

(Class): Apples are somewhat circular, somewhat red, possibly green, and may have stems at the top.



Adaptive Boosting

ML and Life

- combine simple rules to approximate complex function
  (many heads are better than one)
- emphasize incorrect data for valuable information
  (again you can learn by correcting mistakes)

AdaBoost Algorithm

- Input: examples (picture x_n, label $y_n)_{n=1}^{N}$.
- For $t = 1, 2, \cdots, T$,
  - learn a simple rule h_t from emphasized examples
  - get the confidence w_t of such rule
  - emphasize the examples that do not agree with h_t.
- Output: weighted vote of the rules $\sum_{t=1}^{T} w_t h_t(x)$

Machine Learning Research

- What can machines learn? (application)
  - concrete applications (and data mining):
  - abstract setups:
    classification, regression, $\cdots$
- Why can machines learn? (theory)
  - theoretical paradigms:
    statistical learning, reinforcement learning, interactive learning, $\cdots$
  - generalization guarantees
- How can machines learn? (algorithm)
  - faster algorithms
  - algorithms with better **generalization** performance

new opportunities of machine learning keep coming from new applications