Machine Learning

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

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

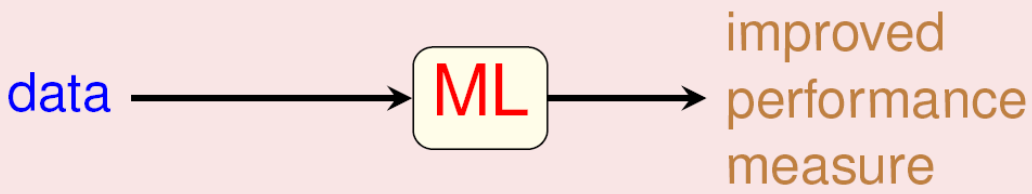
machine learning: acquiring skill
with experience accumulated/**computed** from data

What is skill?

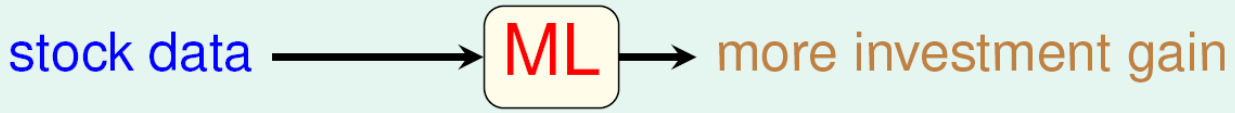
A More Concrete Definition

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?

Dog 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
- 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

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
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

沒有data自然不能ML，例如說要辨識陽性陰性，只有陰(陽)性資料是辦不到的

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 無關
2. determining whether a given graph contains a cycle 現有演算法可解決
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 未被摧毀過 => 無該種資料 => no

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

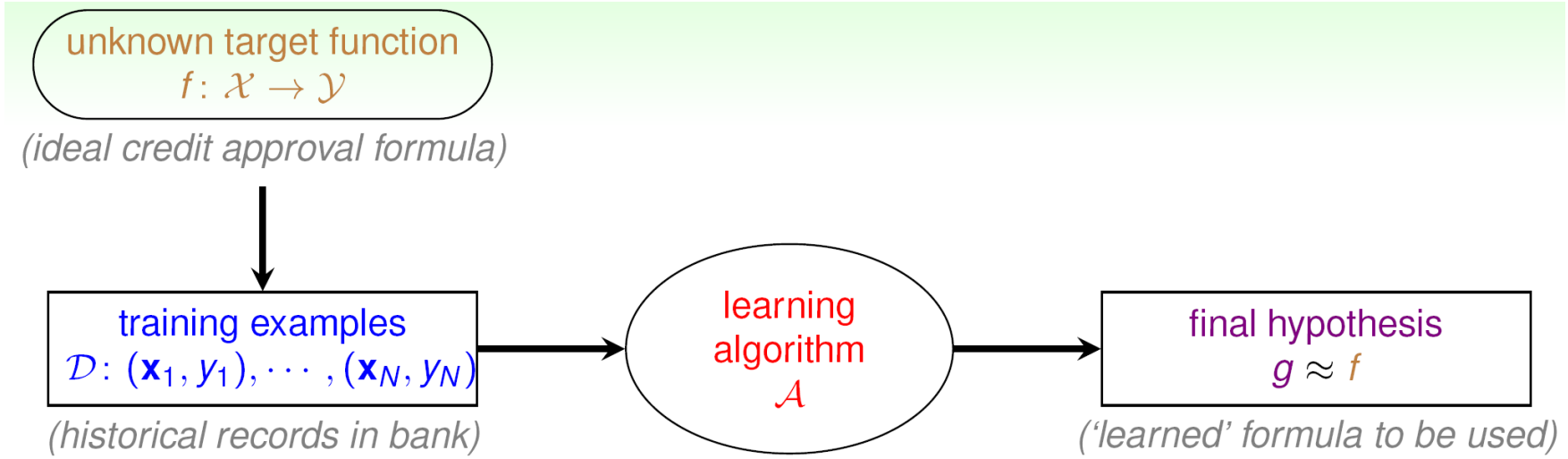
Basic Notations

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

model

$\{(\mathbf{x}_n, y_n)\}$ from f → ML → g

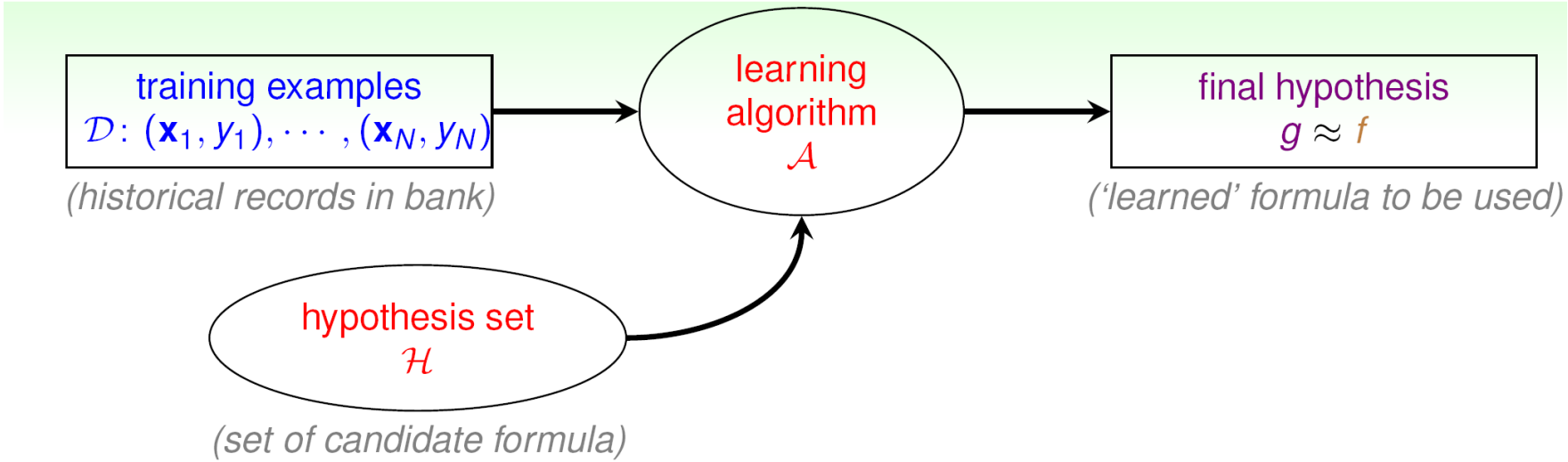
Learning Flow



- 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?

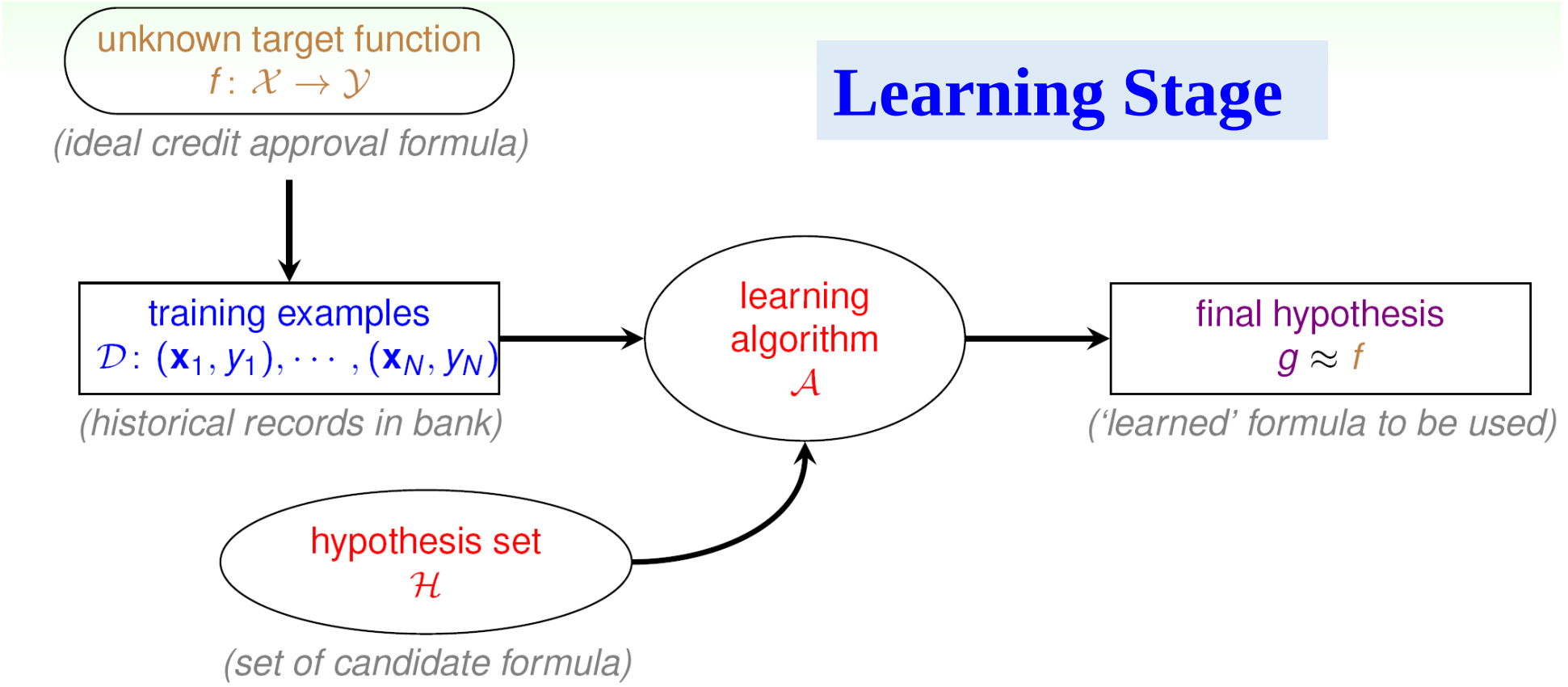
The Learning Model



- 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

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

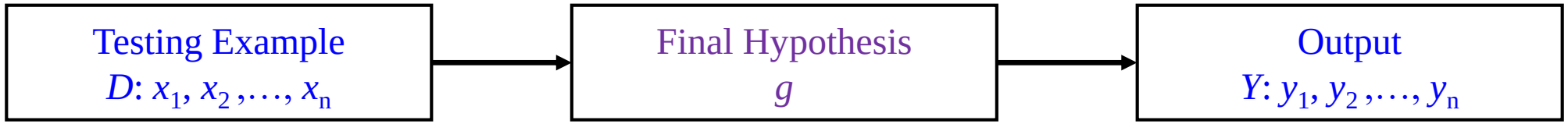
Practical Definition of Machine Learning



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

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

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

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

Machine Learning and Artificial Intelligence

Machine Learning	Artificial Intelligence
use data to compute hypothesis g that approximates target f	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 Data Mining

Machine Learning	Data Mining
use data to compute hypothesis *g* that approximates target *f*	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

Data Mining – 從大量資料中挖掘有趣的特性

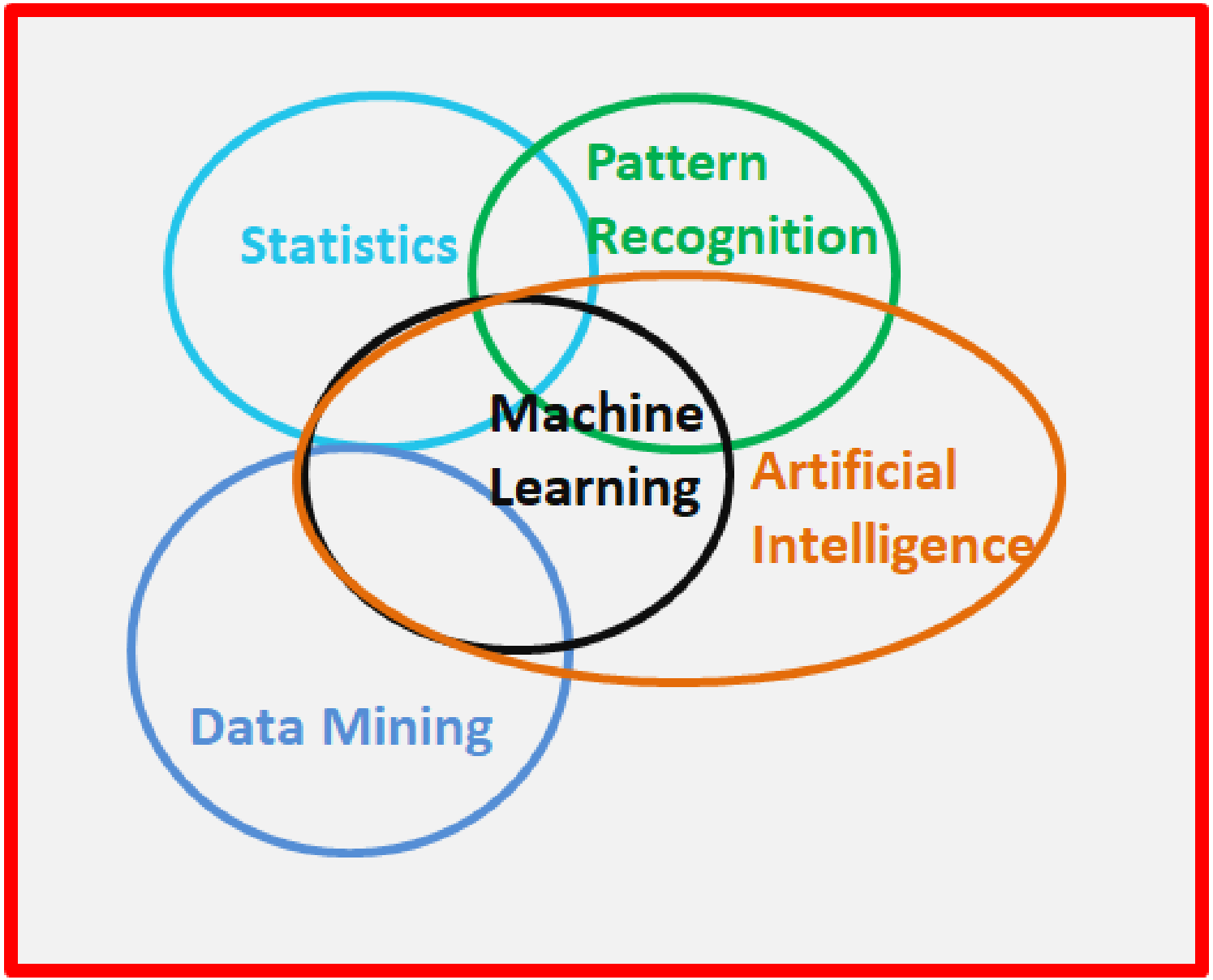
- 英國倫敦基金公司Derwent Capital Markets於2011年曾利用Twitter上發表的推文去統計大眾情緒以預測股市走勢，因此在當年全球市場低迷之時，還能維持1.85%的報酬率，和S&P500下跌了2.2%的指數相比，領先許多。
- 美國零售商Target利用公司內部所擁有的消費者購買資料進行分析，並由此去預測消費者的行為—例如預測孕婦在懷孕初、中期大概會想購買甚麼樣的物品；然後當消費者在網路上購買了某一項產品，系統就會自動提供更多其可能會感興趣的產品資訊。
  - 當時Target寄送了孕婦用品廣告到有可能購買的消費者家中，其中一位收到廣告的孩子父親非常生氣，特地跑到Target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。

Machine Learning and Statistics

Machine Learning	Statistics
use data to compute hypothesis g that approximates target f	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


Statistics
Pattern Recognition
Machine Learning
Artificial Intelligence
Data Mining
DATA SCIENCE

Summary

- 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