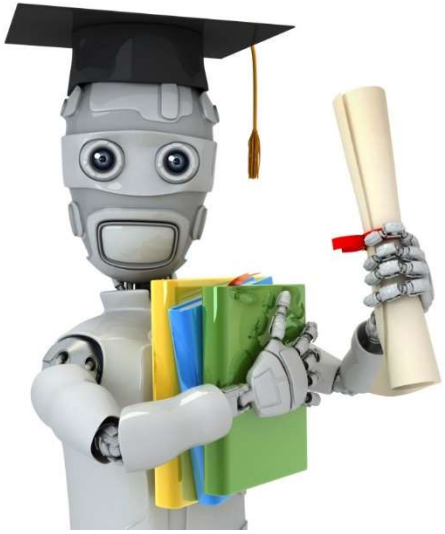


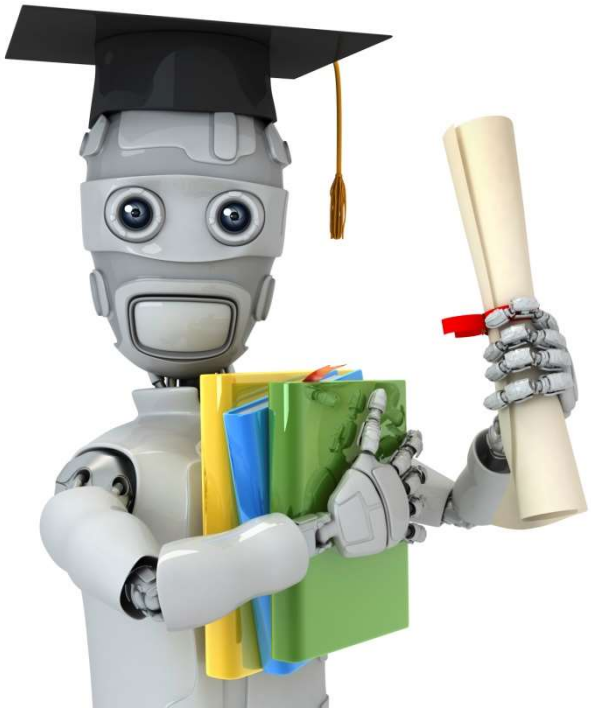
Introduction

Machine Learning

梁毅雄

yxliang@csu.edu.cn

Some materials from Andrew Ng, Hung-yi Lee, Eric Xing, Zico Kolter and others



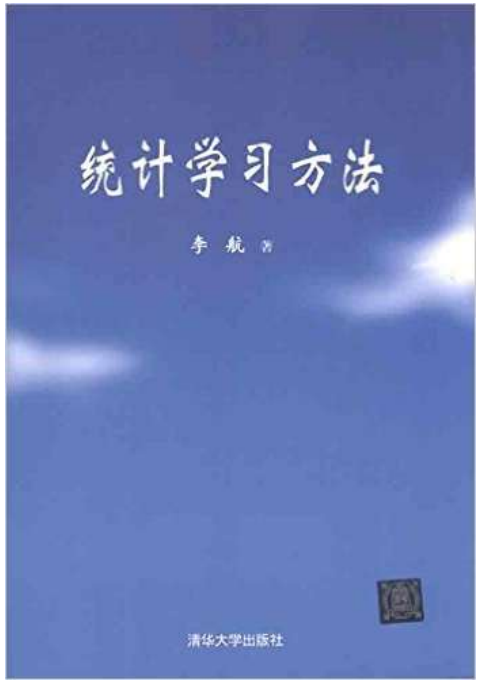
Machine Learning

课程简介

Welcome

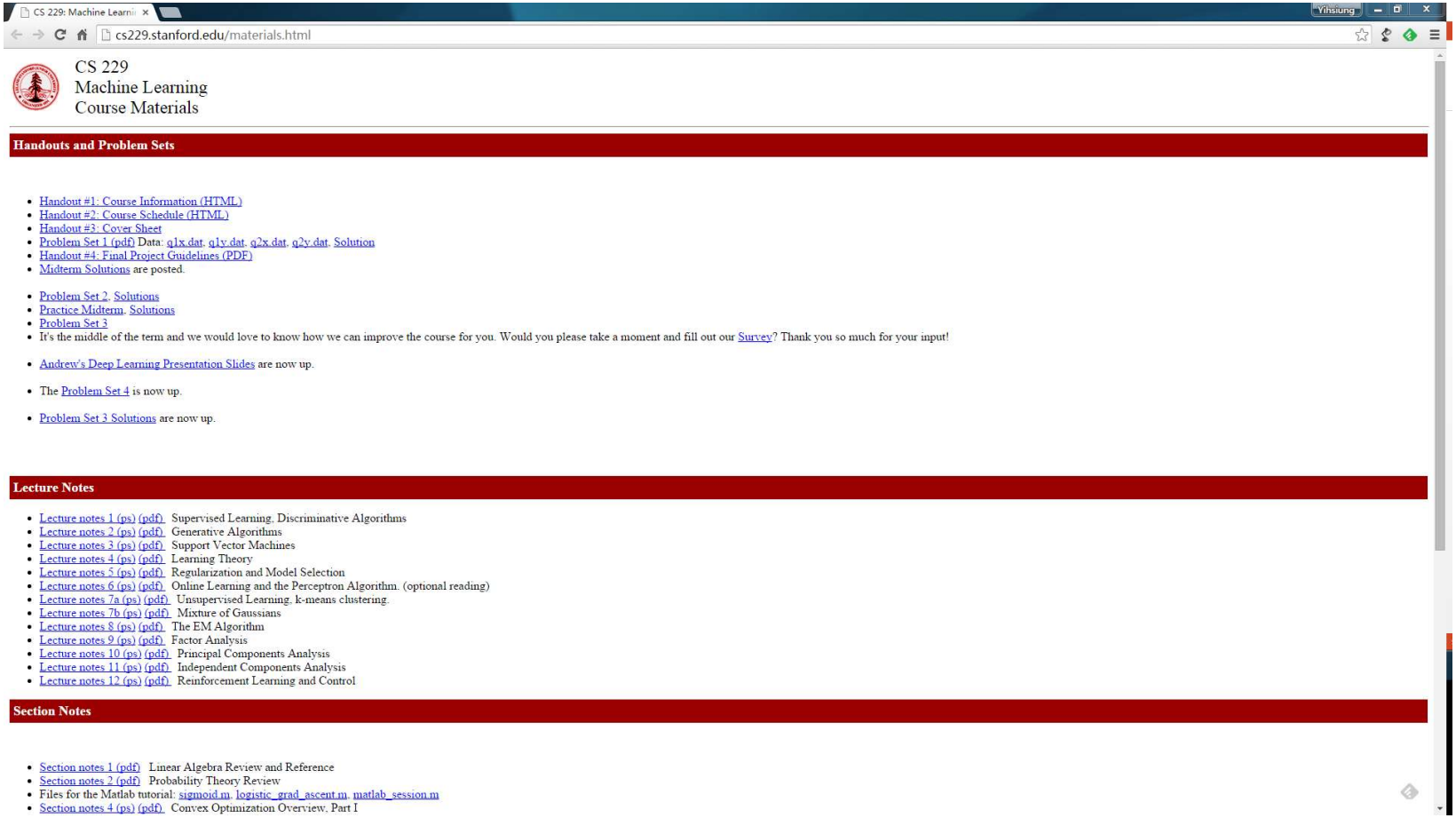
教材与参考资料

- 机器学习
  - 周志华著

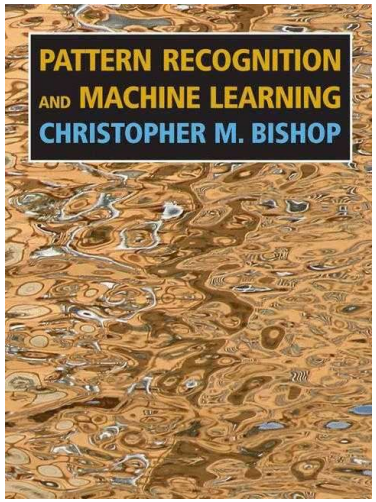


- 统计学习方法
  - 李航著

教材与参考资料


CS 229: Machine Learni
cs229.stanford.edu/materials.html
CS 229
Machine Learning
Course Materials
Handouts and Problem Sets
Handout #1: Course Information (HTML)
Handout #2: Course Schedule (HTML)
Handout #3: Cover Sheet
Problem Set 1 (pdf) Data: q1x.dat, q1y.dat, q2x.dat, q2y.dat, Solution
Handout #4: Final Project Guidelines (PDF)
Midterm Solutions are posted.
Problem Set 2, Solutions
Practice Midterm, Solutions
Problem Set 3
It's the middle of the term and we would love to know how we can improve the course for you. Would you please take a moment and fill out our Survey? Thank you so much for your input!
Andrew's Deep Learning Presentation Slides are now up.
The Problem Set 4 is now up.
Problem Set 3 Solutions are now up.
Lecture Notes
Lecture notes 1 (ps) (pdf) Supervised Learning, Discriminative Algorithms
Lecture notes 2 (ps) (pdf) Generative Algorithms
Lecture notes 3 (ps) (pdf) Support Vector Machines
Lecture notes 4 (ps) (pdf) Learning Theory
Lecture notes 5 (ps) (pdf) Regularization and Model Selection
Lecture notes 6 (ps) (pdf) Online Learning and the Perceptron Algorithm. (optional reading)
Lecture notes 7a (ps) (pdf) Unsupervised Learning, k-means clustering.
Lecture notes 7b (ps) (pdf) Mixture of Gaussians
Lecture notes 8 (ps) (pdf) The EM Algorithm
Lecture notes 9 (ps) (pdf) Factor Analysis
Lecture notes 10 (ps) (pdf) Principal Components Analysis
Lecture notes 11 (ps) (pdf) Independent Components Analysis
Lecture notes 12 (ps) (pdf) Reinforcement Learning and Control
Section Notes
Section notes 1 (pdf) Linear Algebra Review and Reference
Section notes 2 (pdf) Probability Theory Review
Files for the Matlab tutorial: sigmoid.m, logistic_grad_ascent.m, matlab_session.m
Section notes 4 (ps) (pdf) Convex Optimization Overview, Part I

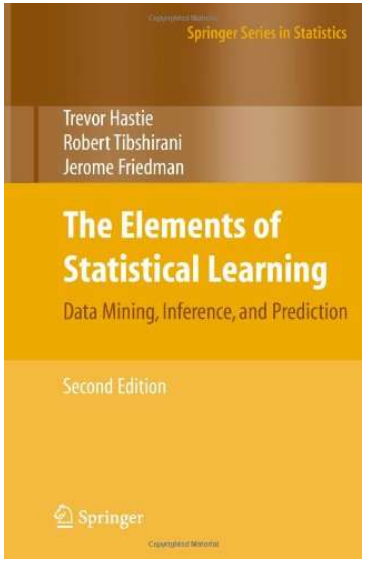
教材与参考资料



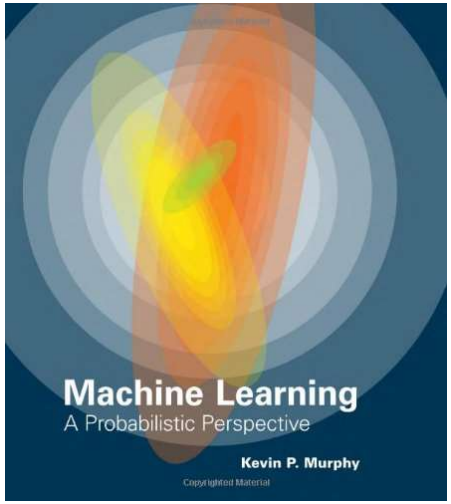
- Pattern Recognition and Machine Learning
  - Christopher Bishop

- 机器学习
  - T Mitchell著, 曾华军等译

教材与参考资料



- The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition
  - Trevor Hastie, Robert Tibshirani, Jerome Friedman



- Machine Learning A Probabilistic Perspective
  - Kevin P. Murphy

教材与参考资料

- www.coursera.org
  - Machine Learning （入门级，本课程）
  - 机器学习基石（Machine Learning Foundations） （初级）
  - 机器学习技法（Machine Learning Techniques） （中级）
- Stanford ML course by Andrew Ng
  - Materials: http://cs229.stanford.edu/materials.html （中级）
- CMU ML course by Tom Mitchell
  - http://www.cs.cmu.edu/~tom/10701_sp11/lectures.shtml （中级）
- CMU ML course by Eric Xing
  - http://www.cs.cmu.edu/~epxing/Class/10701/lecture.html （高级）
- Caltech ML course by Yaser S. Abu-Mostafa
  - https://work.caltech.edu/telecourse.html#lectures （中高级）
- NTU ML course by Hung-yi Lee
  - http://speech.ee.ntu.edu.tw/~tlkagk/courses_ML17_2.html
- UC Berkley, MIT, …

需要具备的基础

- 数学：
  - 应具有较好的数学基础，包括线性代数、概率论、微积分等；
- 编程：
  - 应具有较好的编码基础，至少掌握一门高级语言
  - 作业采用Python语言
- 英语：
  - 具有较好的文献阅读理解能力

考核方式

- 成绩：
  - 平时50%
  - 期末考试 50%（开卷考试）

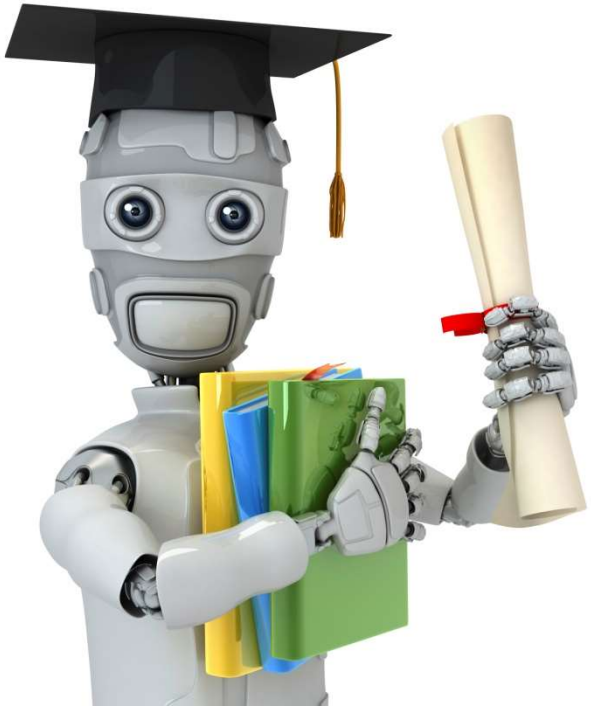
- 平时成绩主要包括：课后作业、平时到课、平时表现
  - 作业请在deadline前提交
  - 作业代码要符合指定格式，经TA修改后方能运行的代码会被扣分

QQ群讨论

- 课程QQ群：722648171
- 仅限于讨论本课程的学习讨论
- 有问题可以直接在群内发问
- 如果有同学知道答案的请帮忙回答
- 好的问题、回答、留言等，平时成绩会加分

Welcome our TAs

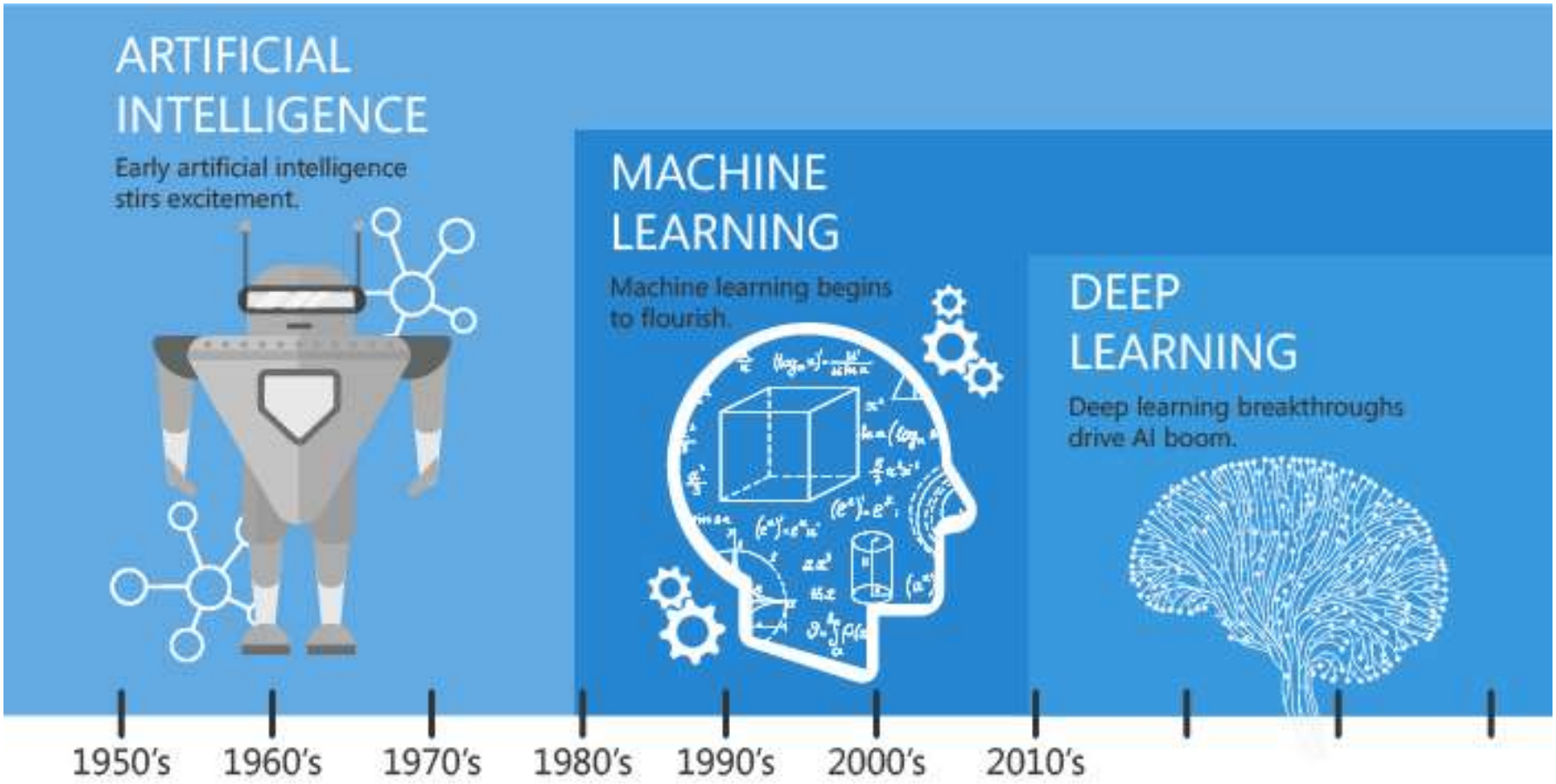
- 赵杨: 1173201671@qq.com)
- 冯硕: 690361877@qq.com
- 王都: 765816845@qq.com
- 赵嘉伟: 229251421@qq.com



Machine Learning

课程简介

概况



Since an early flush of optimism in the 1950's, smaller subsets of artificial intelligence - first machine learning, then deep learning, a subset of machine learning - have created ever larger disruptions.

http://blog.aimagnifi.com/index.php/2017/10/13/what-is-the-difference-between-machine-learning-and-deep-learning/

What is Machine Learning?

This is “cat”

Learning ……

You write the program for learning.

“monkey”

“cat”

“dog”

A large amount of images

[Source: Hung-yi Lee]

机器学习的定义(short)

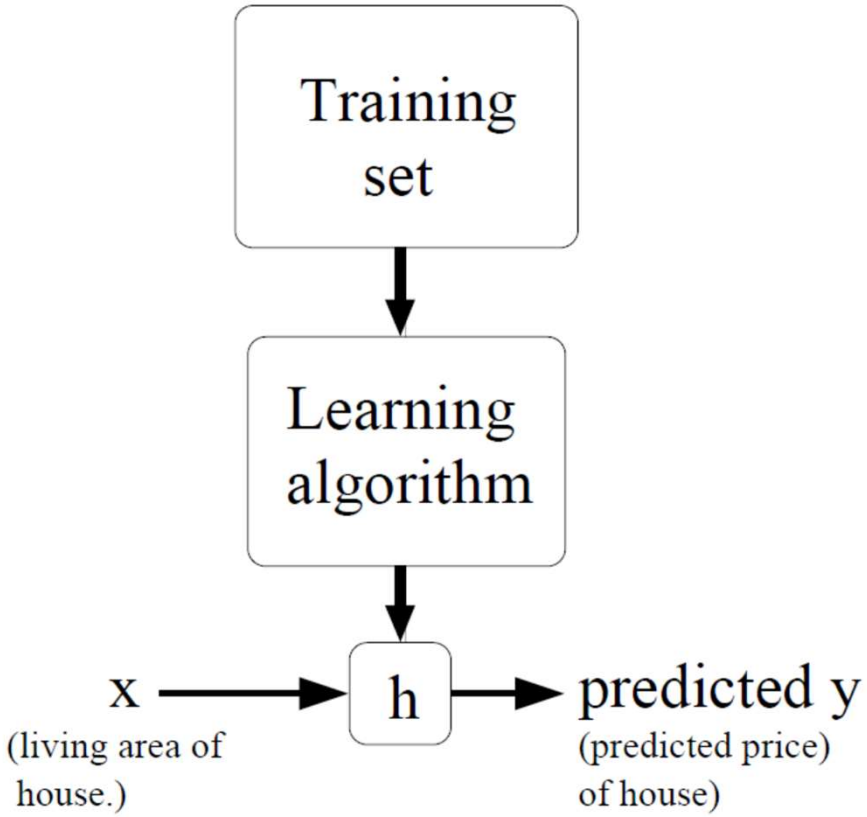
- Arthur Samuel (1959). 机器学习: 它研究的是这样一个学习领域，赋予计算机一种不用显示编程就能够学习的能力。
- Tom Mitchell (1998) 也适当的定义机器学习: 对于某类任务T和性能度量P，如果一个计算机程序在T上以P衡量的性能随着经验E而自我完善，那么我们称这个计算机程序在从经验E学习。

机器学习的定义(Long)

- Machine Learning seeks to develop theories and computer systems for
  - **representing;**
  - **classifying, clustering, recognizing, organizing;**
  - **reasoning under uncertainty;**
  - **predicting;**
  - **…**

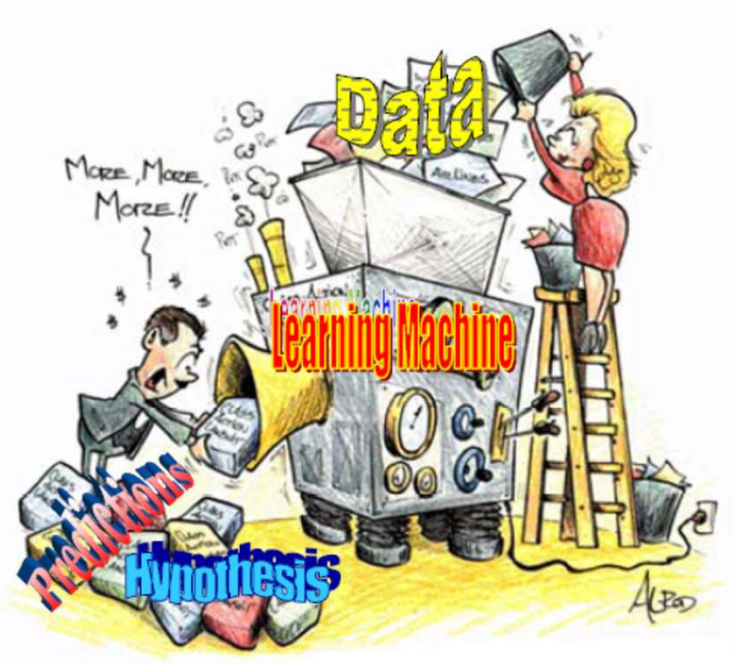
  complex, real world data, based on the system's own experience with data, and (hopefully) under a unified model or mathematical framework, that
    - **can be formally characterized and analyzed**
    - **can take into account human prior knowledge**
    - **can generalize and adapt across data and domains**
    - **can operate automatically and autonomously**
    - **and can be interpreted and perceived by human**

机器学习框架



h: a hypothesis function

机器学习框架

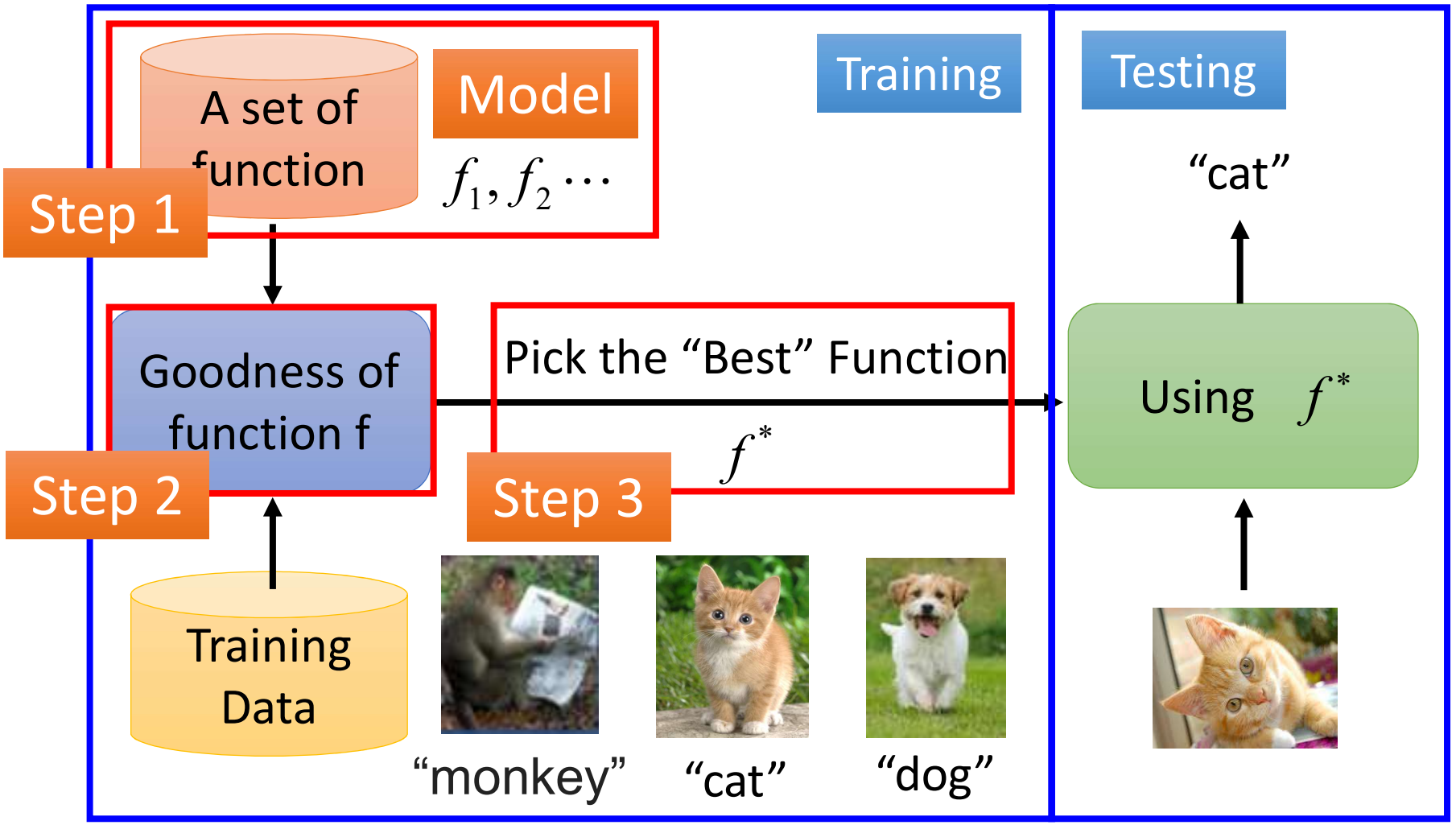


- 将任务表示为输入到输出的映射
- 将性能表示为损失或目标函数
- Examples:
  - Medical Diagnosis
    - mapping input to one of several classes/categories Classification
  - Predict tomorrow's Temperature
    - mapping input to a number Regression
  - Chance of Survival: From patient data to p(survive >= 5 years)
    - mapping input to probability Logistic Regression
  - Driving recommendation
    - mapping input into a plan Planning

机器学习框架

Image Recognition:

$f($  $)=$ “cat”

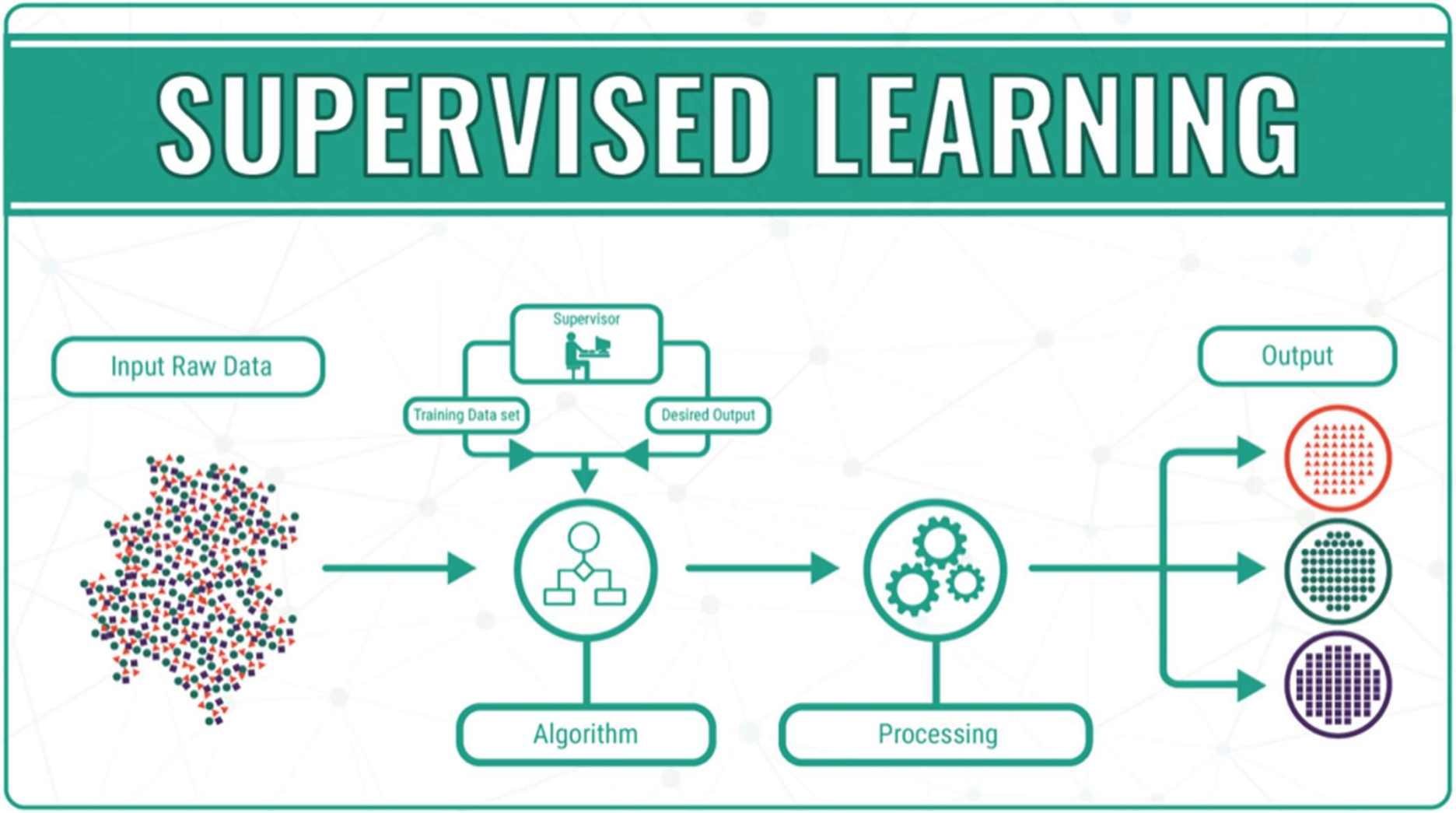


[Source: Hung-yi Lee]

Types of learning problems

- 监督学习（Supervised Learning）
- 非监督学习（Unsupervised Learning）
- 半监督学习（Semi-supervised Learning）
- 迁移学习（Transfer Learning）
- 主动学习（ Active learning）
- 强化学习（Reinforcement Learning）
- 元学习（Meta Learning/Learning to Learn）
- …

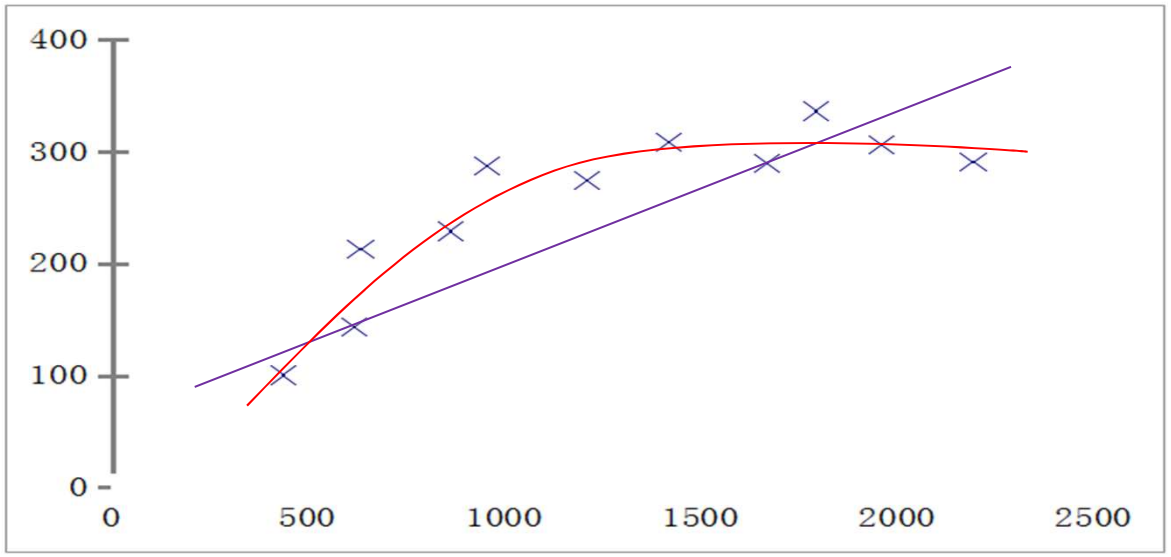
监督学习



http://bigdata-madesimple.com/machine-learning-explained-understanding-supervised-unsupervised-and-reinforcement-learning/

监督学习

Regression：房价预测



监督学习

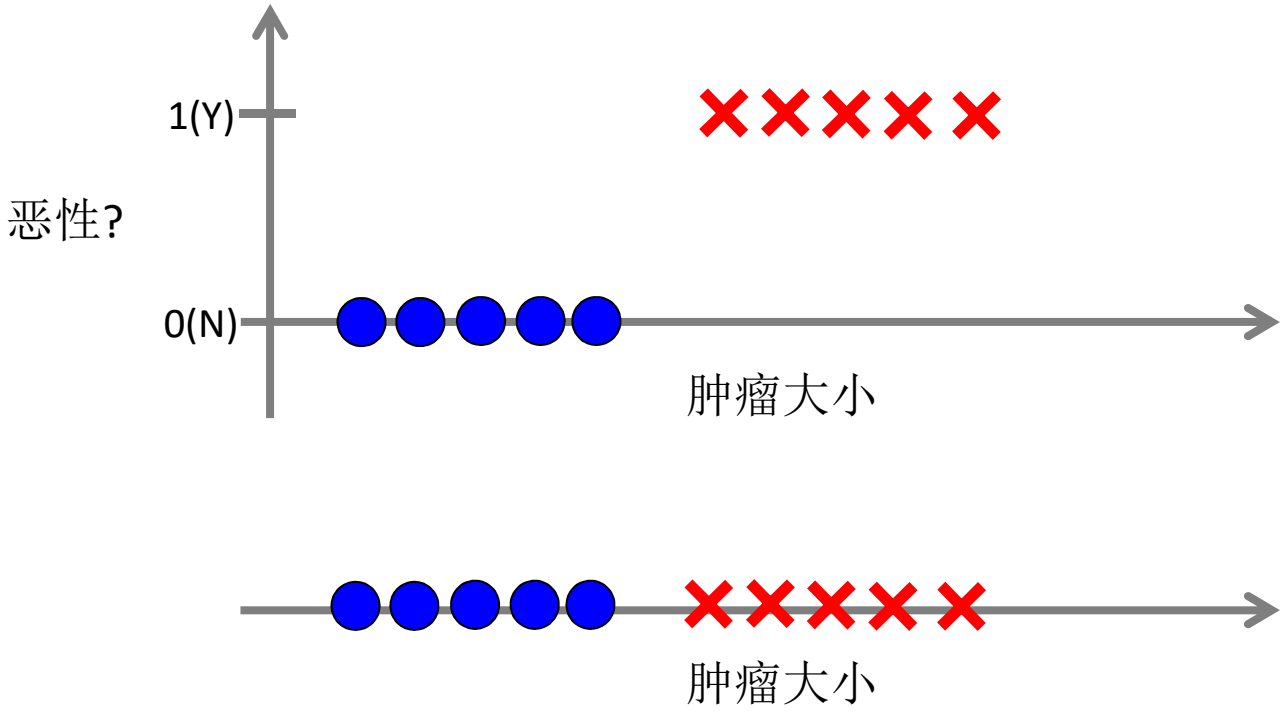
给出“正确答案”

回归算法: 预测连续值输出

(价格)

监督学习

Binary Classification： 癌症诊断(恶性, 良性)

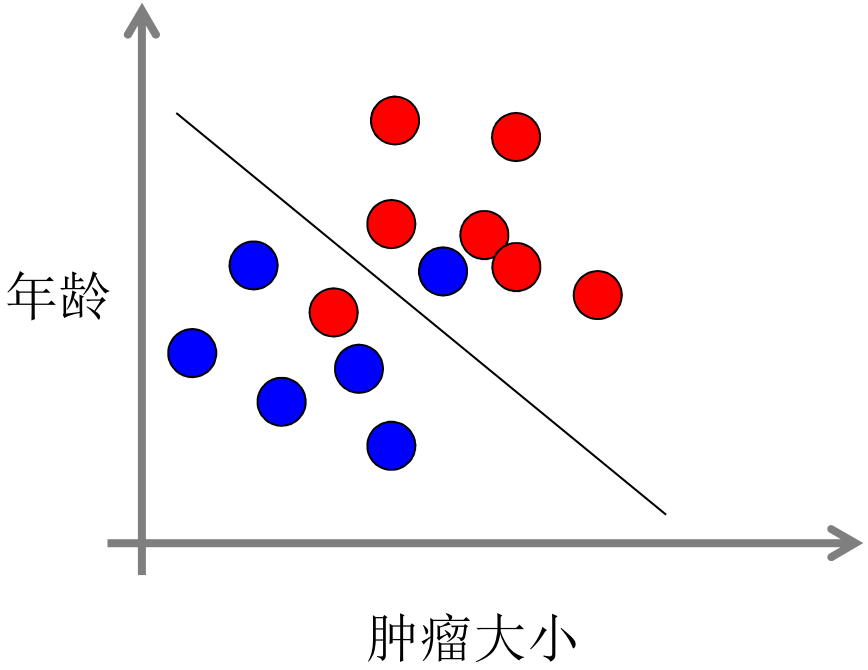


分类

离散值输出 (0 或 1)

监督学习

Binary Classification： 癌症诊断(恶性, 良性)

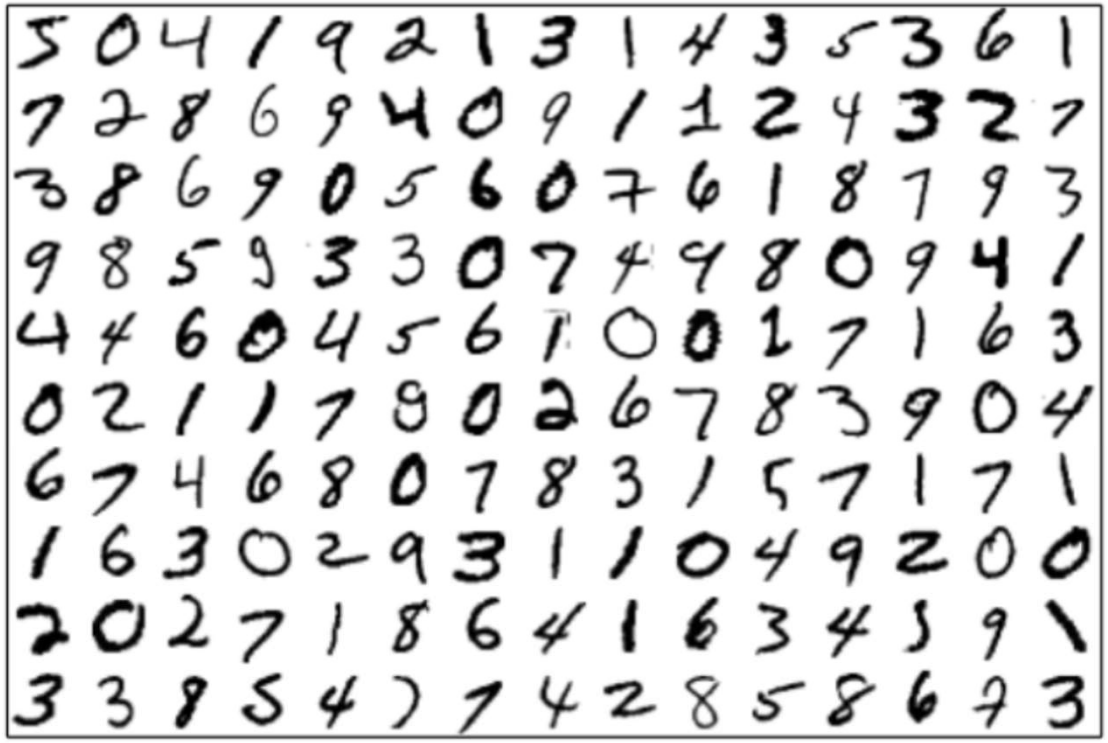


- 肿块密度
- 细胞大小的均匀性
- 细胞形状的均匀性

...

监督学习

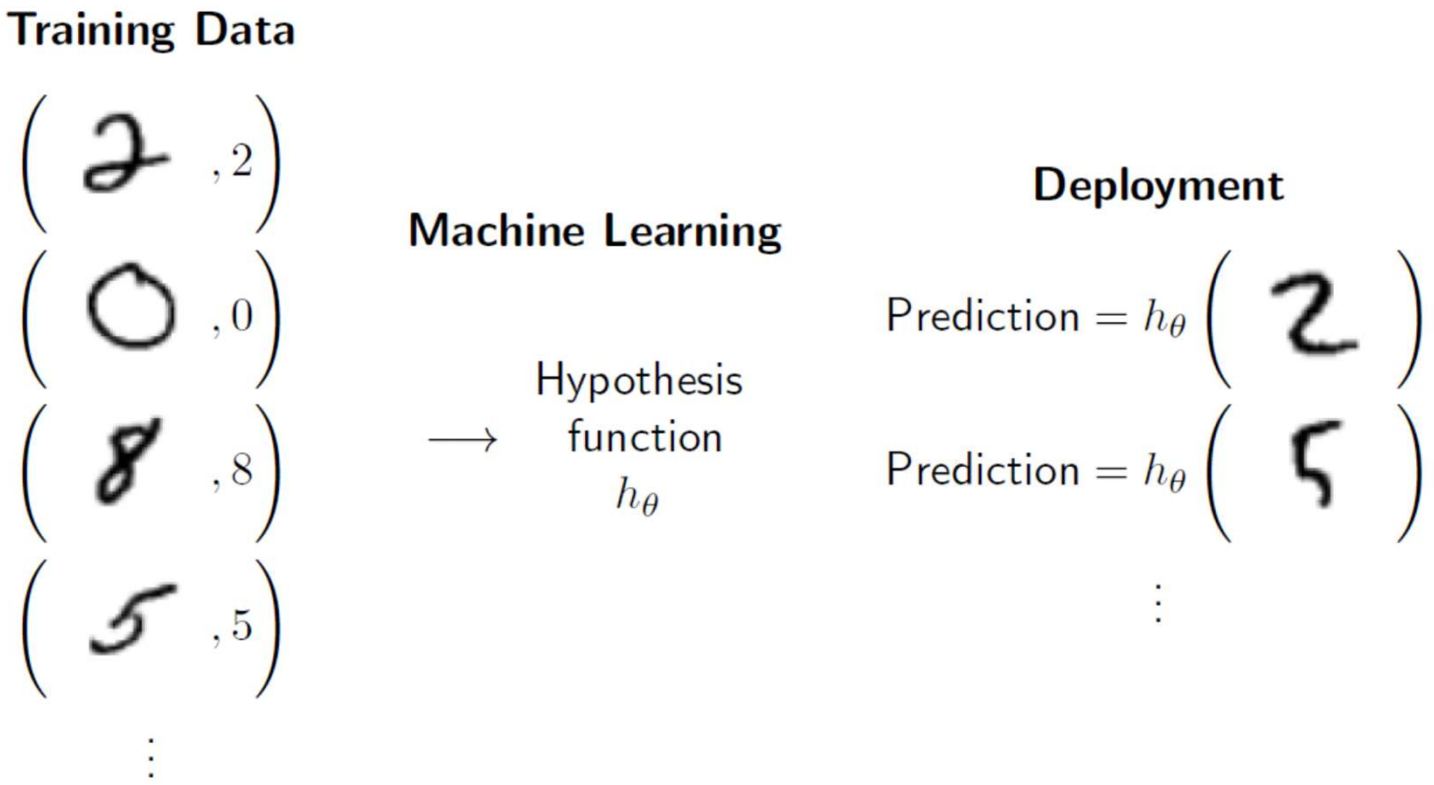
Multiclass Classification: 数字识别



Digits from MNIST dataset
(http://yann.lecun.com/exdb/mnist/)

监督学习

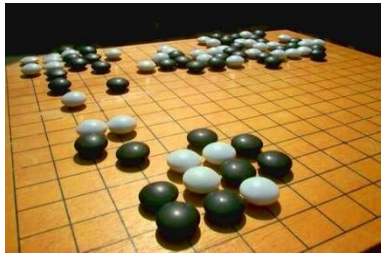
Multiclass Classification: 数字识别



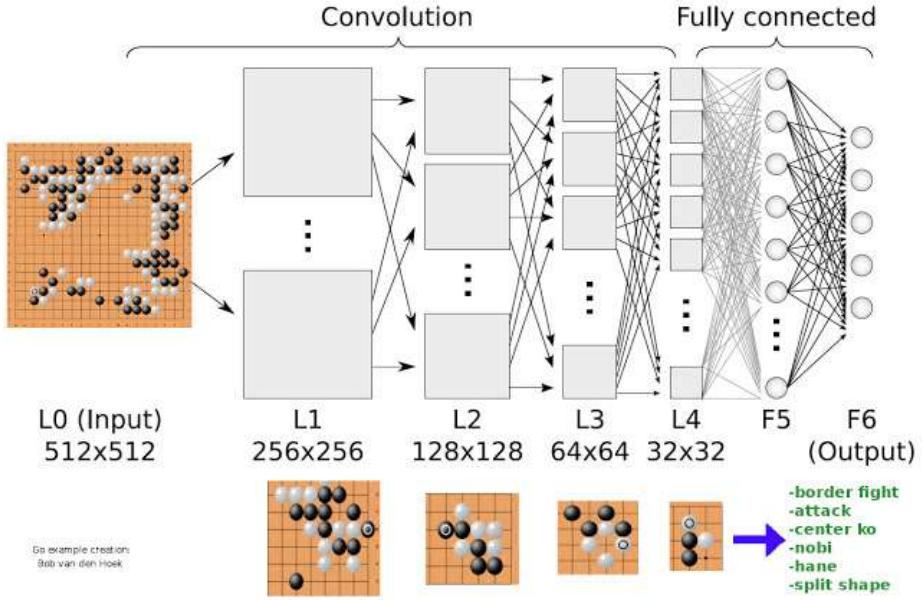
监督学习

- Learn to play Go

Next move:
“5-5”



Next move:
“3-3”



Next move
Each position
is a class
(19 x 19 classes)

http://deeplearningskysthelimit.blogspot.com/2016/04/part-2-alphago-under-magnifying-glass.html

监督学习

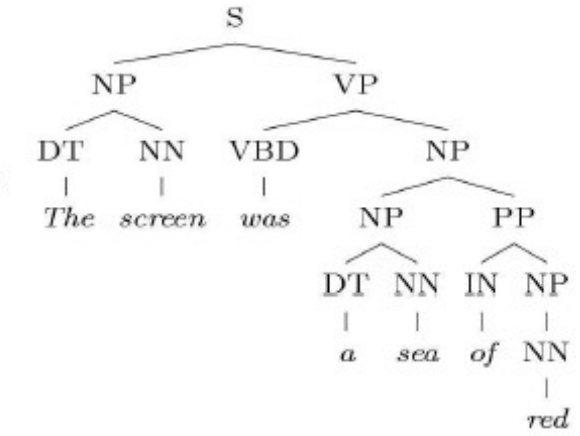
Structured Prediction

brace

Image

Word

The screen was
a sea of red

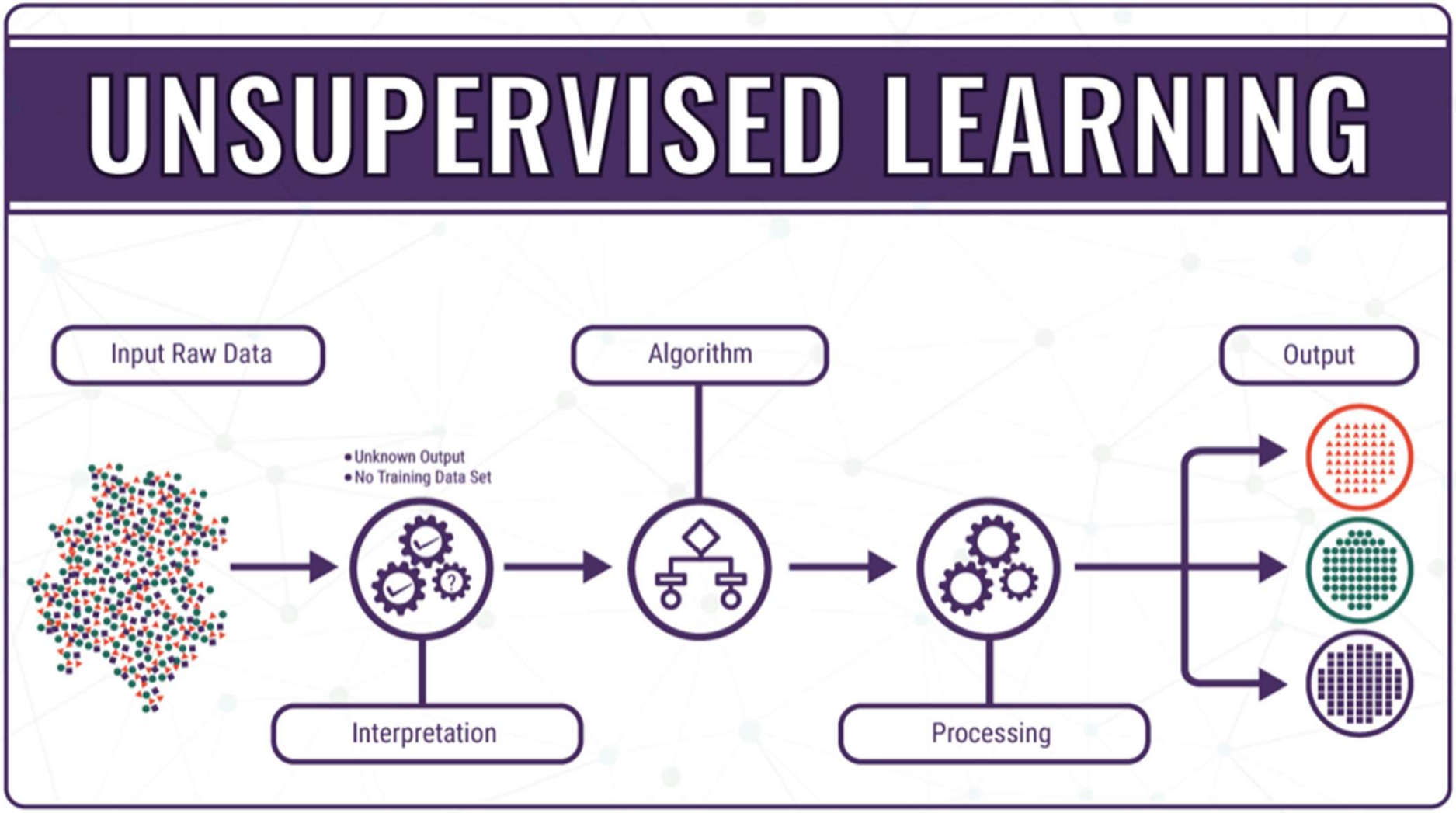


Sentence

Parse tree

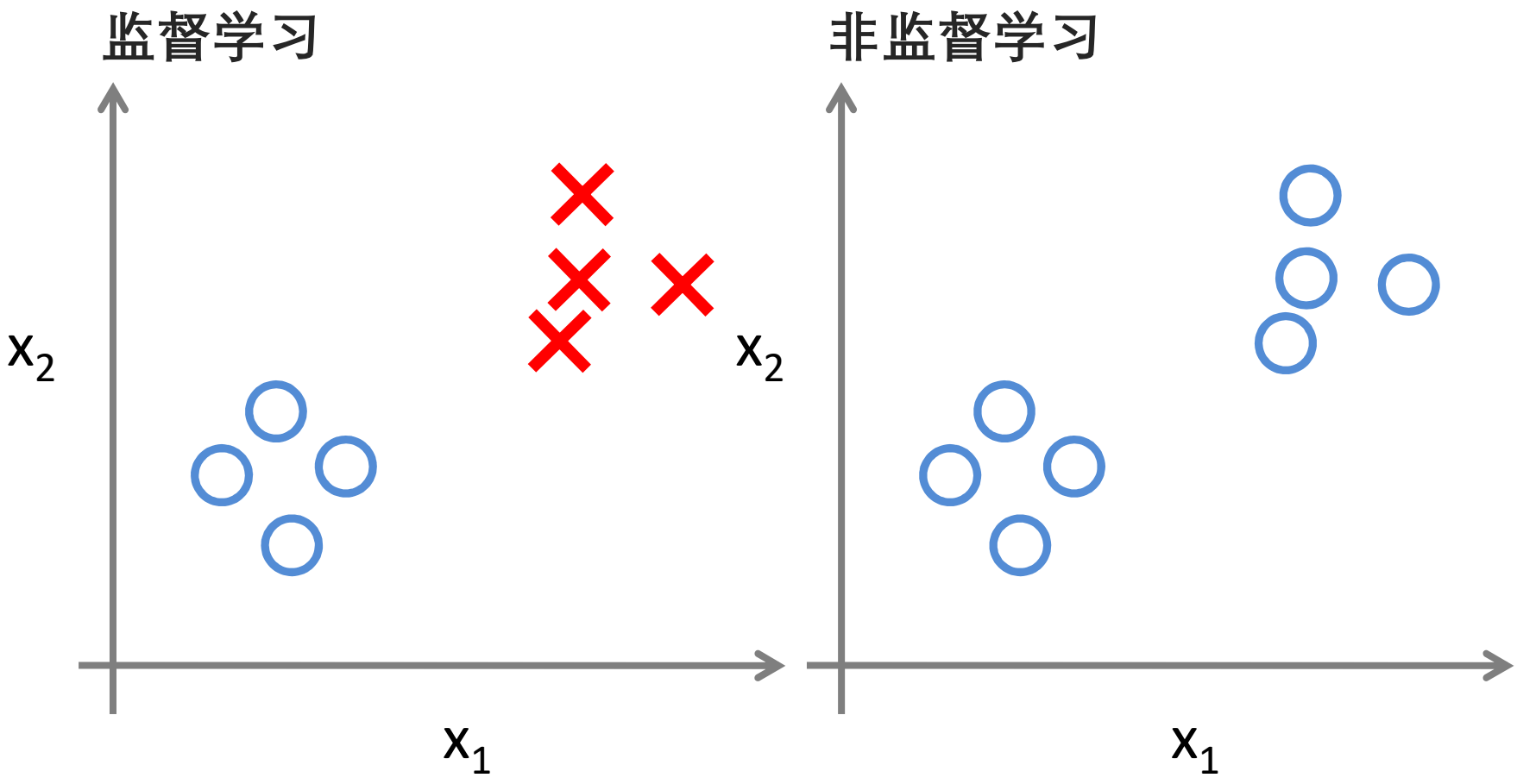
Source: B. Taskar

非监督学习

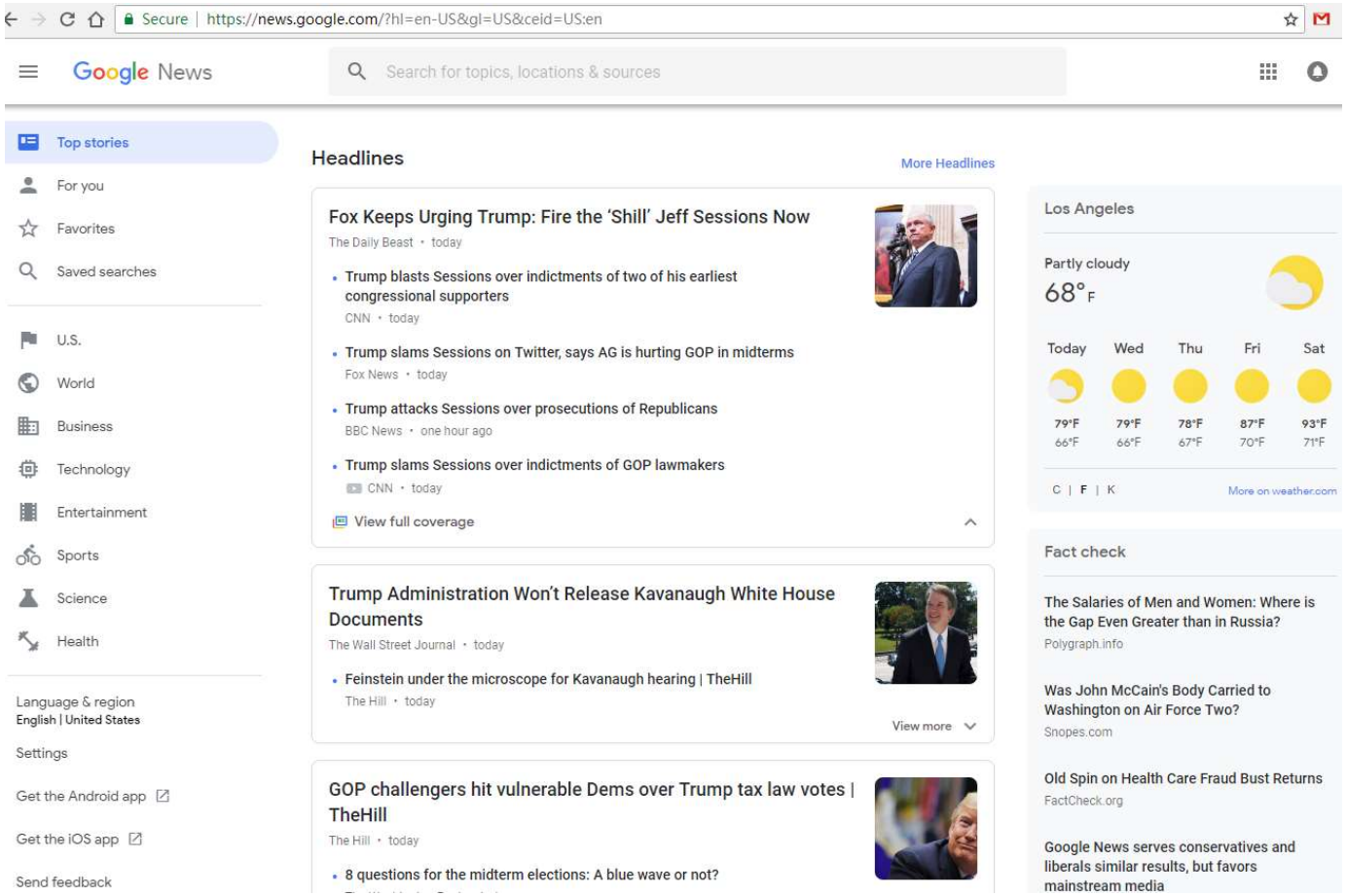


http://bigdata-madesimple.com/machine-learning-explained-understanding-supervised-unsupervised-and-reinforcement-learning/

非监督学习

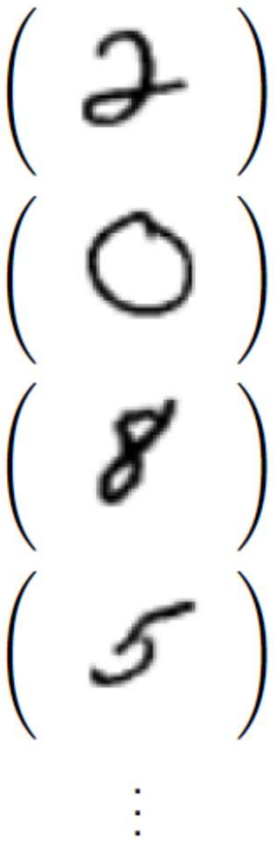


非监督学习


Secure | https://news.google.com/?hl=en-US&gl=US&ceid=US:en
Google News
Search for topics, locations & sources
Top stories
For you
Favorites
Saved searches
U.S.
World
Business
Technology
Entertainment
Sports
Science
Health
Language & region
English | United States
Settings
Get the Android app
Get the iOS app
Send feedback
Headlines
More Headlines
Fox Keeps Urging Trump: Fire the 'Shill' Jeff Sessions Now
The Daily Beast • today
Trump blasts Sessions over indictments of two of his earliest congressional supporters
CNN • today
Trump slams Sessions on Twitter, says AG is hurting GOP in midterms
Fox News • today
Trump attacks Sessions over prosecutions of Republicans
BBC News • one hour ago
Trump slams Sessions over indictments of GOP lawmakers
CNN • today
View full coverage
Trump Administration Won't Release Kavanaugh White House Documents
The Wall Street Journal • today
Feinstein under the microscope for Kavanaugh hearing | TheHill
The Hill • today
View more
GOP challengers hit vulnerable Dems over Trump tax law votes | TheHill
The Hill • today
8 questions for the midterm elections: A blue wave or not?
Los Angeles
Partly cloudy
68° F
Today Wed Thu Fri Sat
79°F 66°F
79°F 66°F
78°F 67°F
87°F 70°F
93°F 71°F
C | F | K
More on weather.com
Fact check
The Salaries of Men and Women: Where is the Gap Even Greater than in Russia?
Polygraph.info
Was John McCain's Body Carried to Washington on Air Force Two?
Snopes.com
Old Spin on Health Care Fraud Bust Returns
FactCheck.org
Google News serves conservatives and liberals similar results, but favors mainstream media

非监督学习

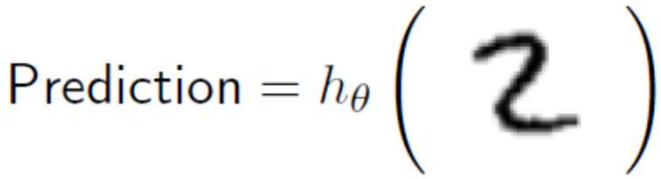
Training Data

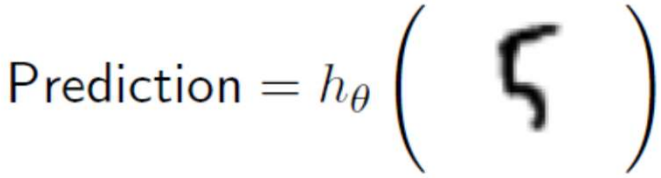


Machine Learning

$\longrightarrow$ Hypothesis function h_θ

Deployment

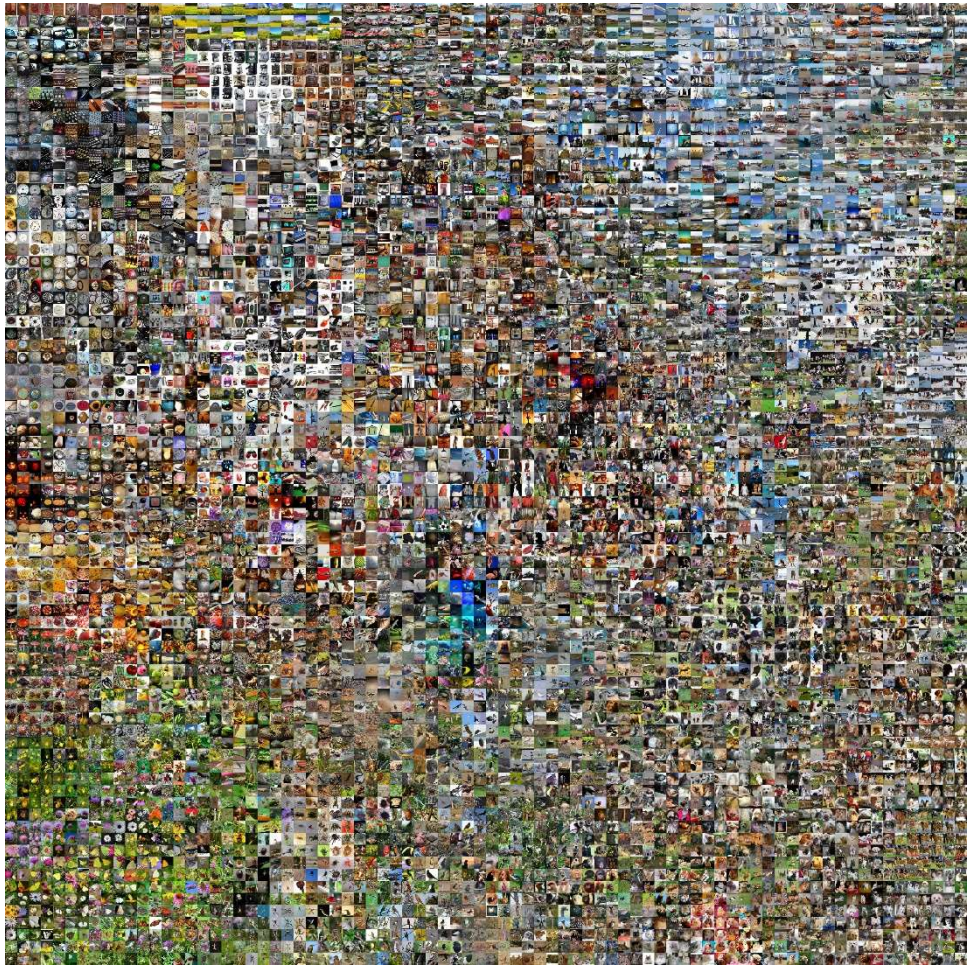
Prediction $= h_\theta$ 

Prediction $= h_\theta$ 

$\vdots$

非监督学习

Maaten, Laurens van der, and Geoffrey Hinton. "Visualizing data using t-SNE." *Journal of machine learning research* 9, no. Nov (2008): 2579-2605.



https://cs.stanford.edu/people/karpathy/cnnembed/

半监督学习

- 半监督学习:
  - Attempts to use unlabeled data as well as labeled data to improve performance
  - Because unlabeled data is often plentiful and labeling data can be expensive
  - Leveraging the large pool of unlabeled examples would be very attractive

半监督学习

For example, recognizing cats and dogs

Labelled data

cat

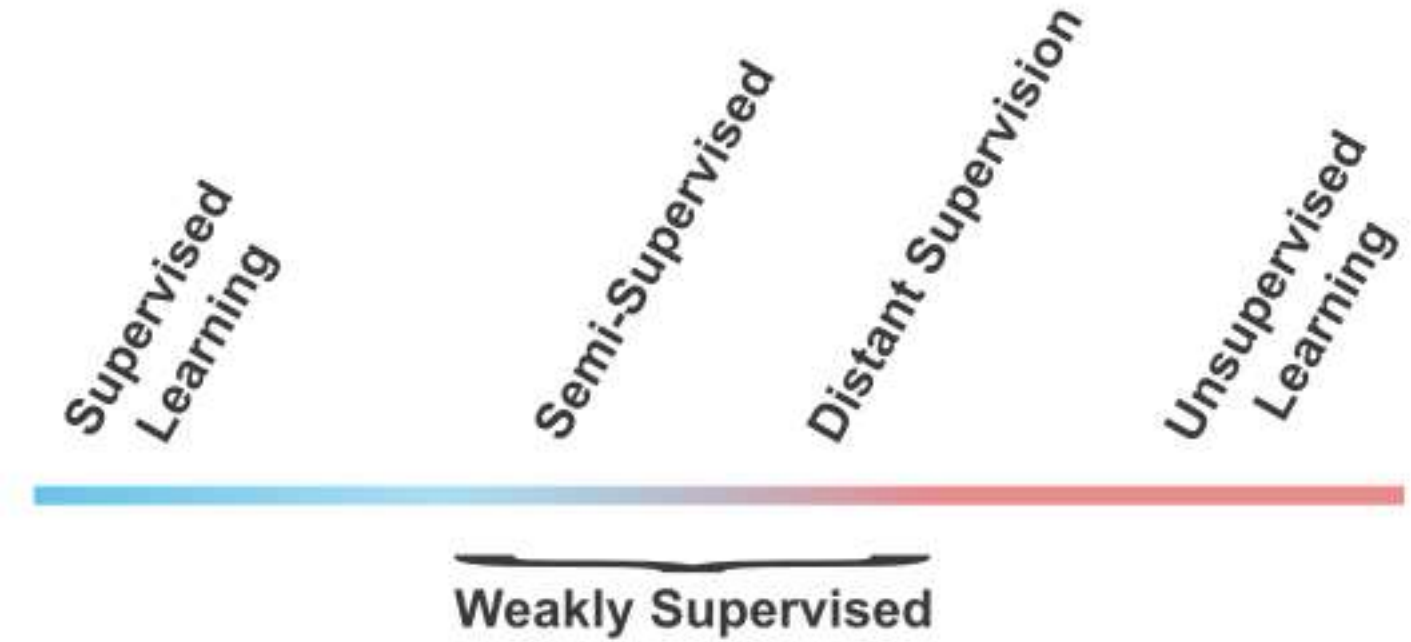
dog

Unlabeled data

(Images of cats and dogs)

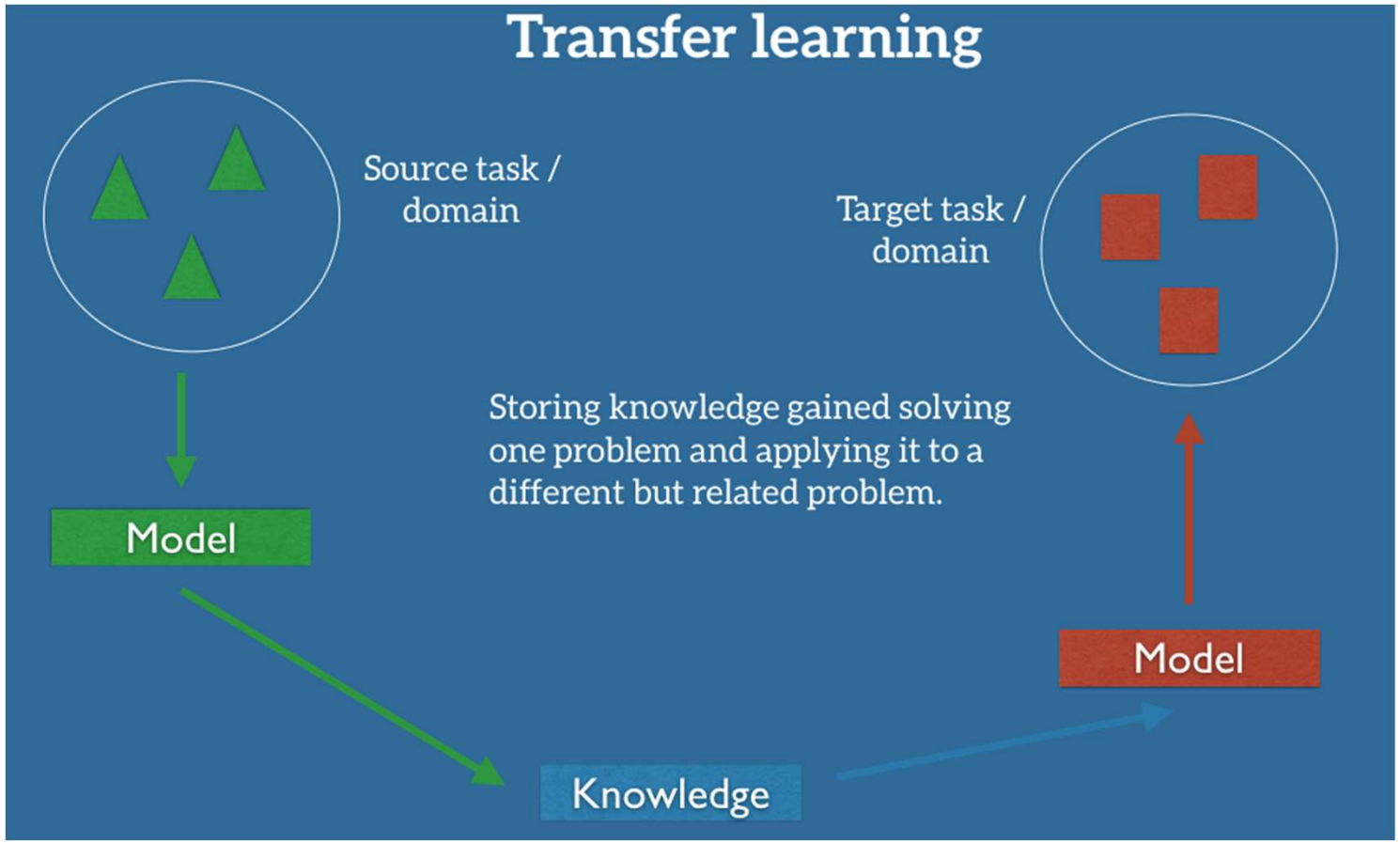
[Source: Hung-yi Lee]

弱监督学习（Weakly Supervised Learning）



http://www.mkbergman.com/1872/knowledge-supervision-as-a-grounding-for-machine-learning/

迁移学习

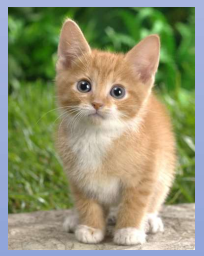


http://ruder.io/transfer-learning/

迁移学习

For example, recognizing cats and dogs

Labelled data

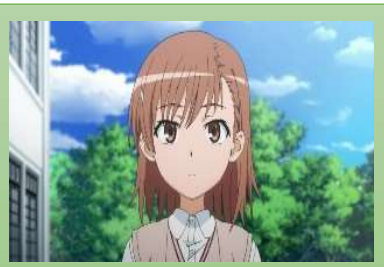


cat

dog

elephant

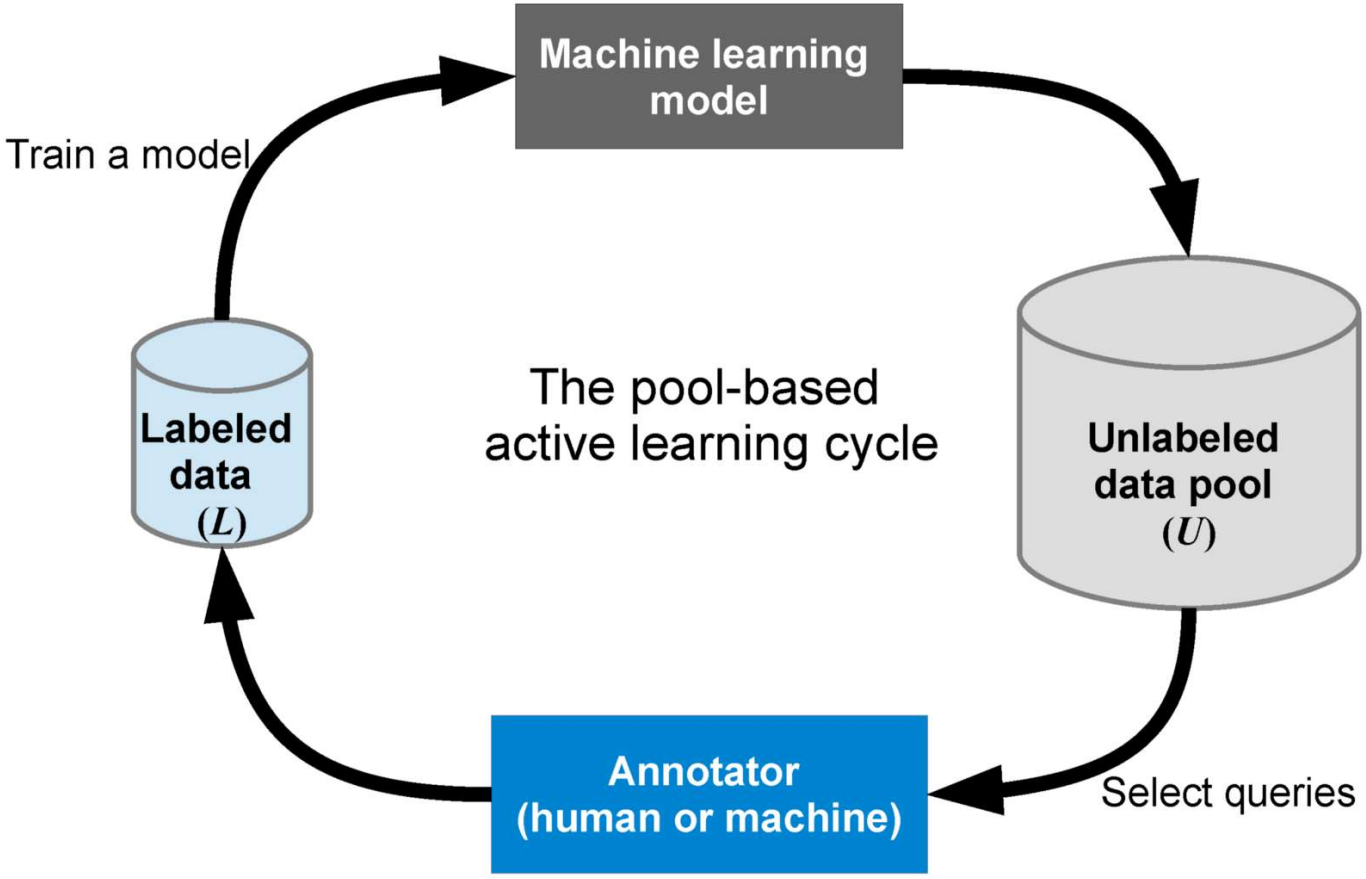
Haruhi



Data not related to the task considered (can be either labeled or unlabeled)

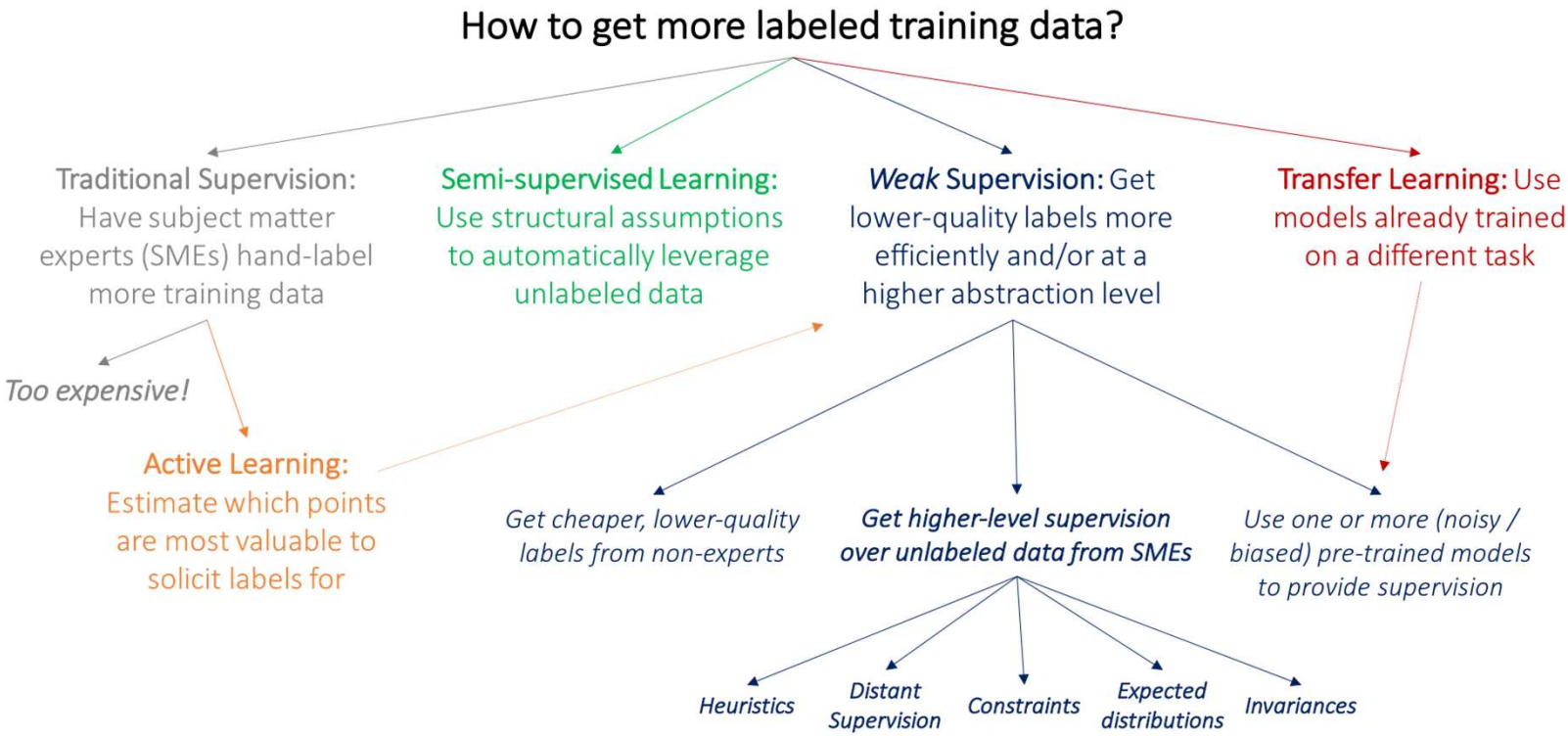
[Source: Hung-yi Lee]

主动学习



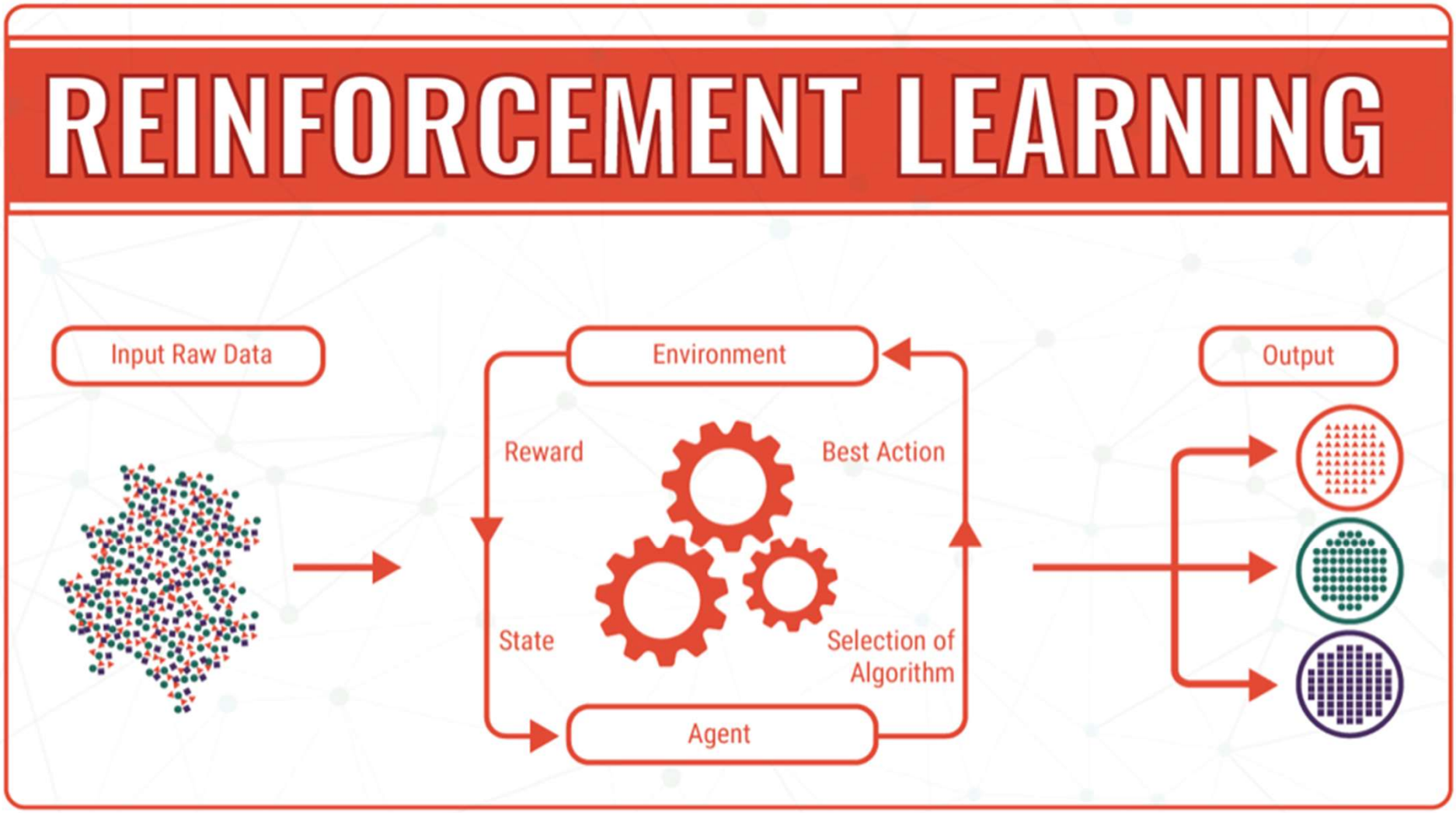
https://medium.com/@kaleajit27/apaperaday-week1-a-meta-learning-approach-to-one-step-active-learning-5ffea59099a2

Data Is at the Heart of the Matter



https://dawn.cs.stanford.edu/2017/07/16/weak-supervision/

强化学习



http://bigdata-madesimple.com/machine-learning-explained-understanding-supervised-unsupervised-and-reinforcement-learning/

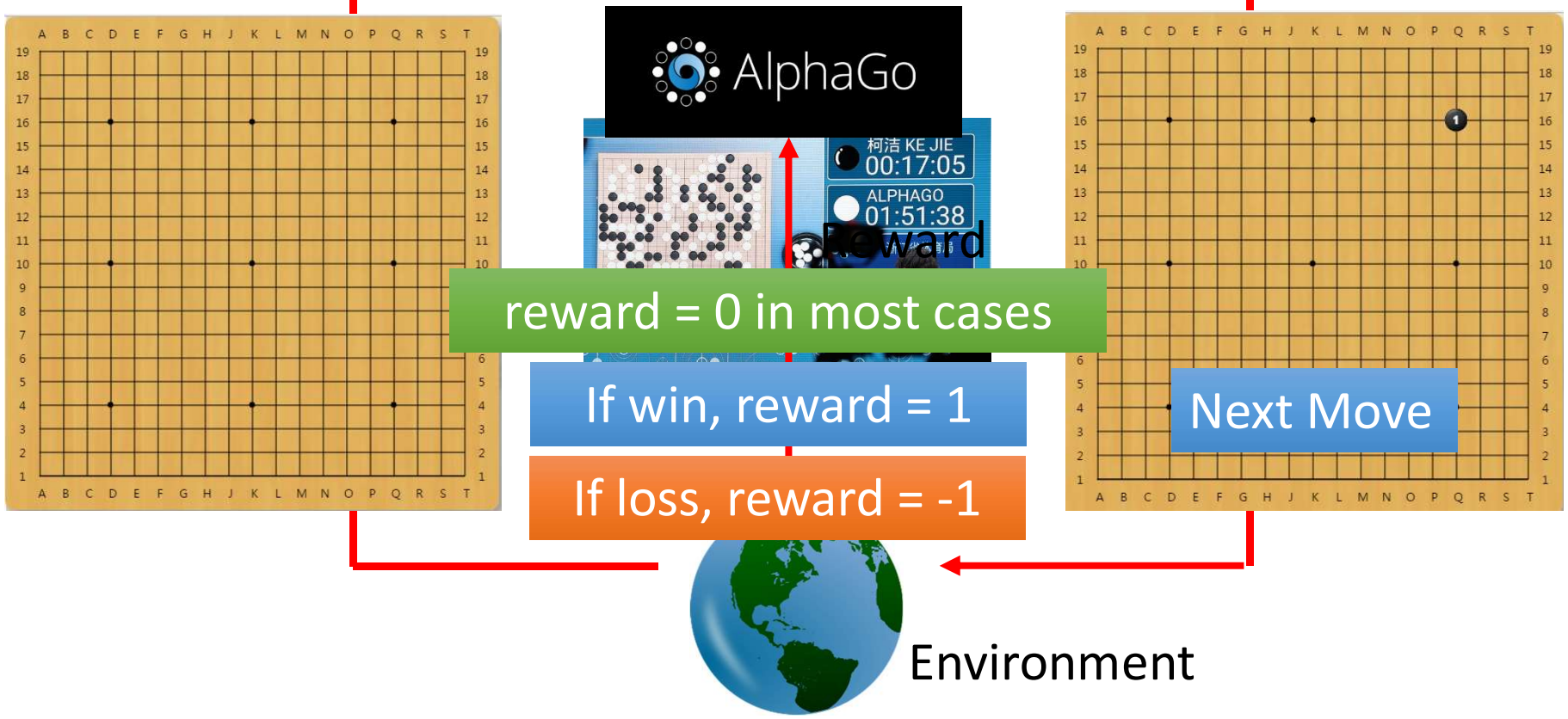
强化学习


nature
LEARNING CURVE
Self-taught AI software
SHARE DATA IN OUTBREAKS
A GIANT IN THE EARLY UNIVERSE
TELEPORTATION FOR TWO


nature
THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE
At last — a computer program that can beat a champion Go player PAGE 484
ALL SYSTEMS GO
CONSERVATION
SONGBIRDS A LA CARTE
Illegal harvest of millions of Mediterranean birds
PAGE 452
RESEARCH ETHICS
SAFEGUARD TRANSPARENCY
Don't let openness backfire on individuals
PAGE 459
POPULAR SCIENCE
WHEN GENES GOT 'SELFISH'
Dawkins's calling card 40 years on
PAGE 462
NATUREASIA.COM
28 January 2016
Vol. 529 No. 7587

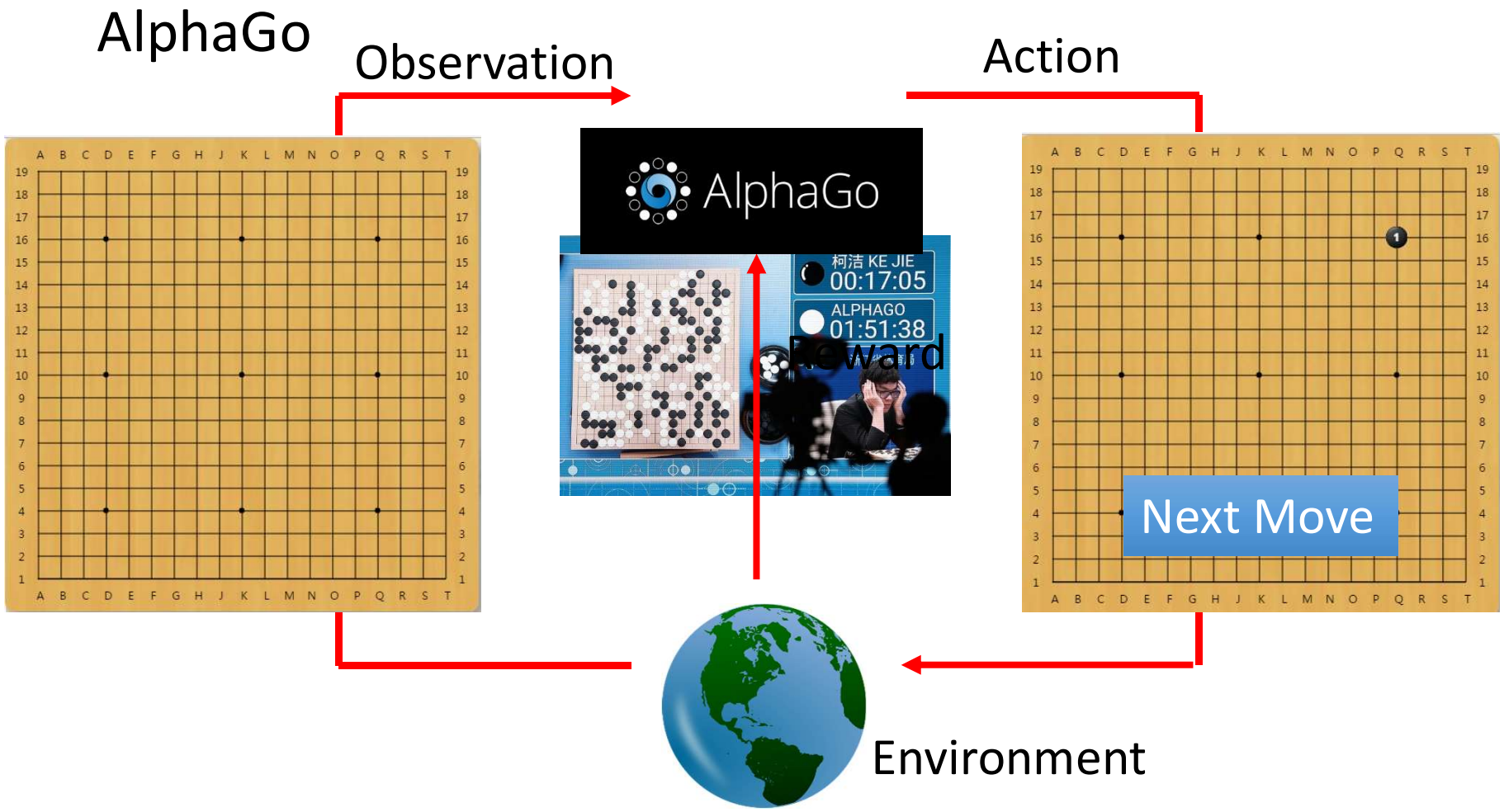
强化学习

Agent learns to take actions maximizing expected reward.



[Source: Hung-yi Lee]

强化学习

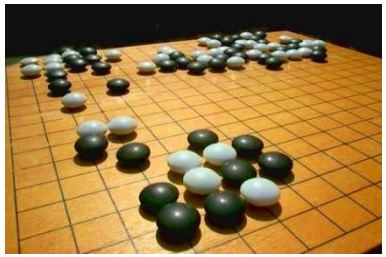


[Source: Hung-yi Lee]

Supervised v.s. Reinforcement

- Supervised: Learning from teacher

Next move: "5-5"



Next move: "3-3"

- Reinforcement Learning Learning from experience

First move → …… many moves …… → Win!

Alpha Go is supervised learning + reinforcement learning.

[Source: Hung-yi Lee]

Thanks!

Any questions?