

Last Chapter

- Bayesian networks provide a natural representation for (causally induced) conditional independence
- Topology + CPTs = compact representation of joint distribution
- Generally easy for domain experts to construct
- Exact inference by variable elimination:
 - ▣ polytime on polytrees, NP-hard on general graphs
 - ▣ space = time, very sensitive to topology
- Naïve Bayes model



Learning from Observations

Chapter 18

Outline

3

- Introduction to machine learning
- Supervised learning (监督学习)
 - ▣ Decision tree learning (决策树学习)
 - ▣ Linear predictions (线性预测)
 - ▣ Support vector machines (支持向量机)
 - ▣ Neural networks (神经网络)
 - ...
- Unsupervised learning (无监督学习)

Learning

4

Learning is essential for unknown environments,

- ▣ i.e., when designer lacks omniscience (全知)

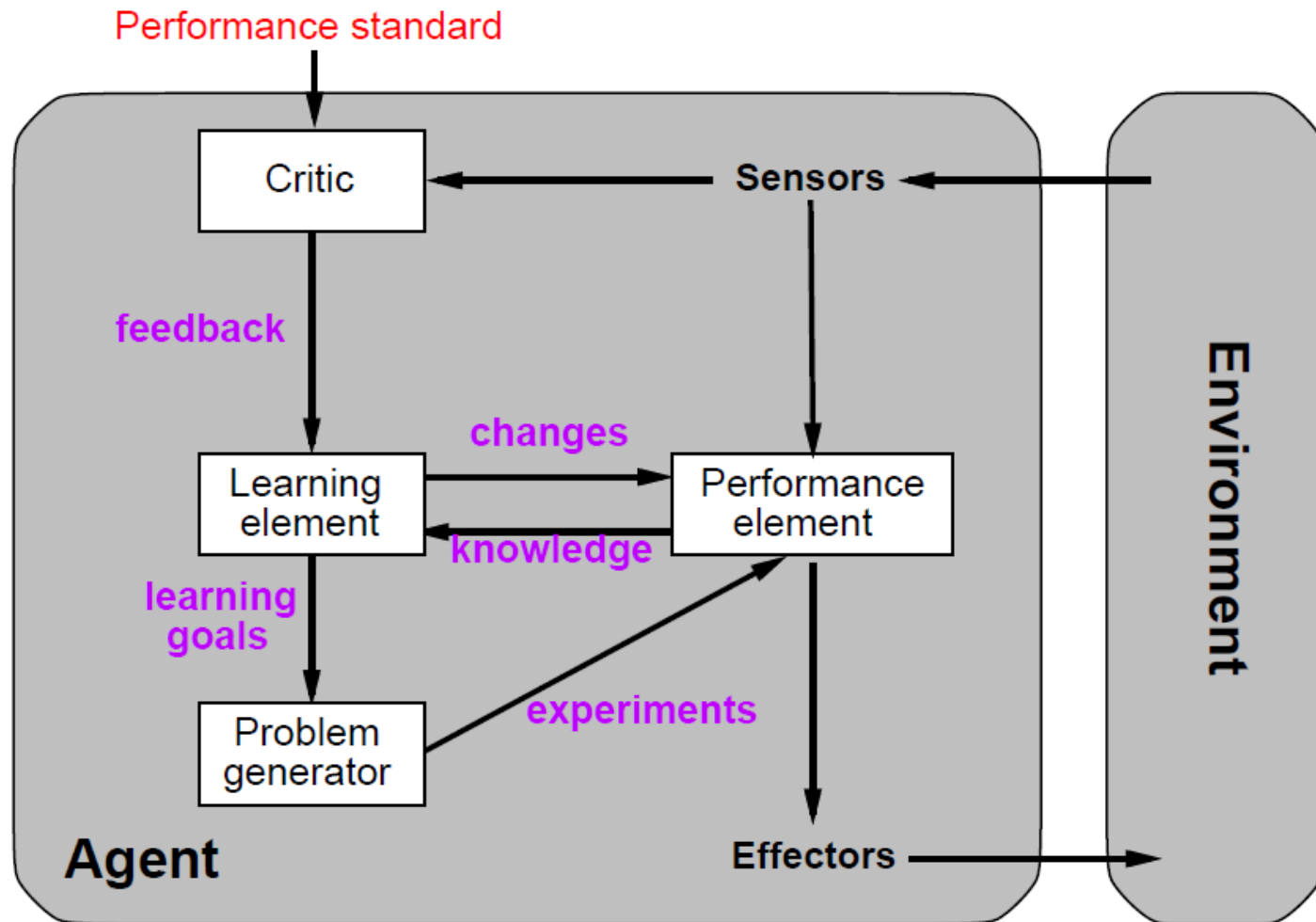
Learning is useful as a system construction method,

- ▣ i.e., expose the agent to reality rather than trying to write it down

Learning modifies the agent's decision mechanisms to improve performance

Learning agents

5



Learning element

6

Design of a learning element is affected by

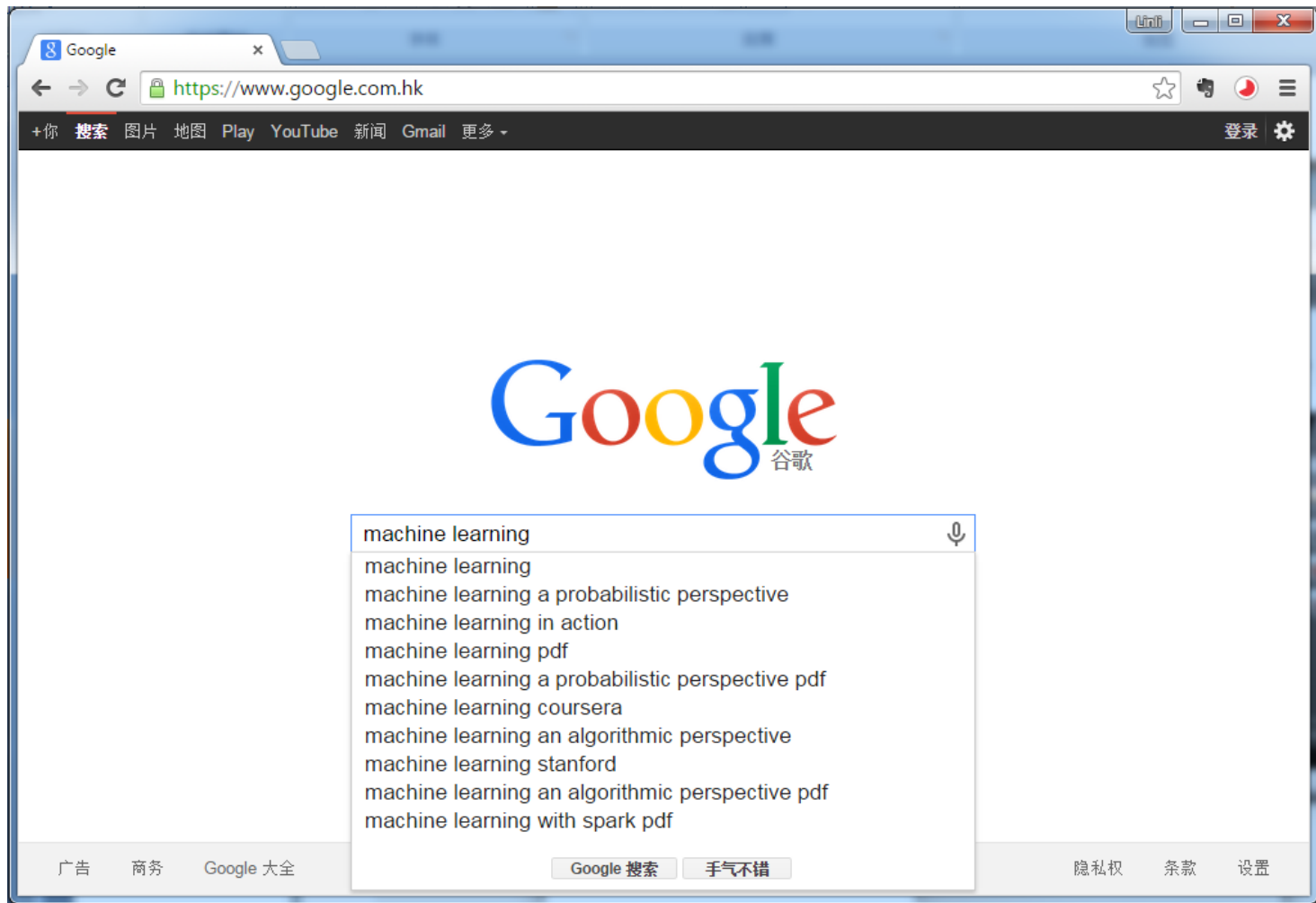
- ▣ Which components of the performance element are to be learned
- ▣ What feedback is available to learn these components
- ▣ What representation is used for the components

7

Introduction to Machine Learning

Machine Learning Everyday: Search Engine

8



Machine Learning Everyday:

Spam Detection (垃圾邮件检测)

9



Machine Learning Everyday:

Machine Translation

10

The spirit is willing but the flesh is weak. [*Bible, Matthew 26:41*]

Дух охотно готов но плоть слаба	Spirit is willingly ready but flesh it is weak
精神是愿意的但骨肉是微弱的	The spirit is wants but the flesh and blood is weak
精神は喜んでであるが、肉は弱い	Mind is rejoicing,, but the meat is weak
El alcohol está dispuesto pero la carne es débil	The alcohol is arranged but the meat is weak
الكحول مستعدة غير أنّ اللحم ضعيف	The alcohol is ready nevertheless the meat is weak.

Statistical machine translation models

Machine Learning Everyday: Face Detection

11

- Now in most digital cameras for auto focusing



Also blink and smile
detection!



Machine Learning

12

- Grew out of work in Artificial Intelligence
- New capability for computers

Why Machine Learning?

13

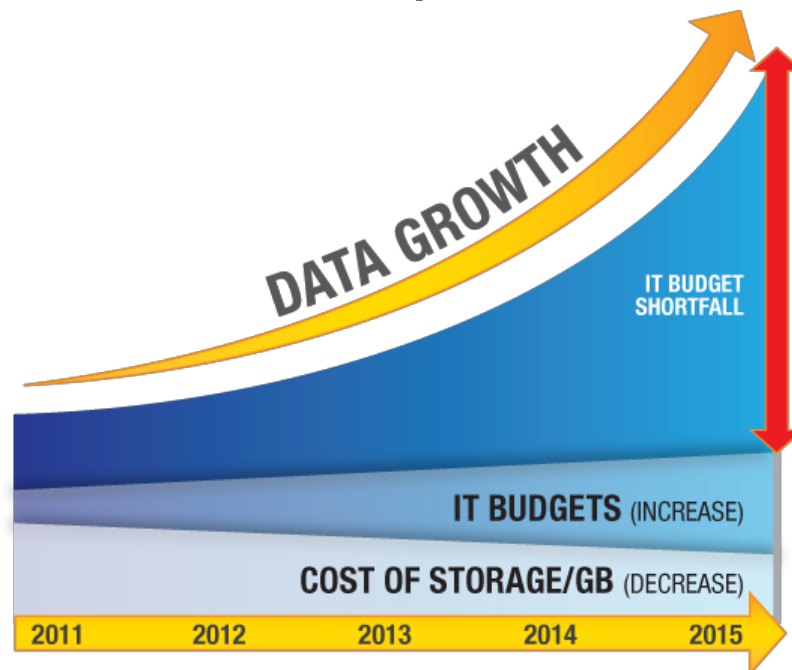
- Solve classification problems
- Learn models of data (“data fitting”)
- Understand and improve efficiency of human learning
- Discover new things or structures that are unknown to humans (“data mining”)

...

Why Machine Learning?

14

- Large amounts of data
 - ▣ Web data, Medical data, Biological data...
- Expensive to analyze by hand
- Computers become cheaper and more powerful

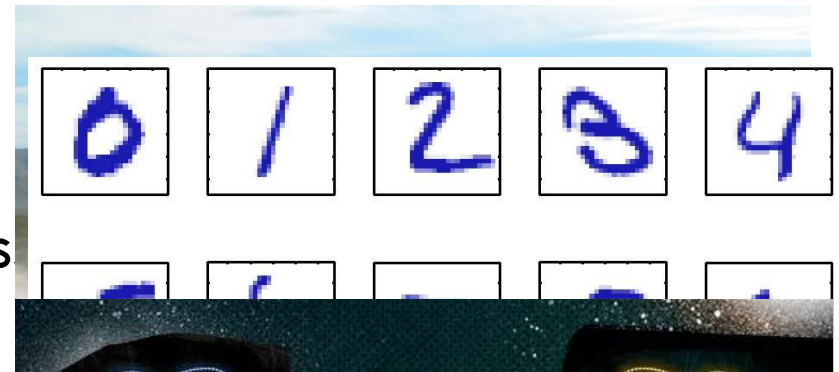


Why Machine Learning?

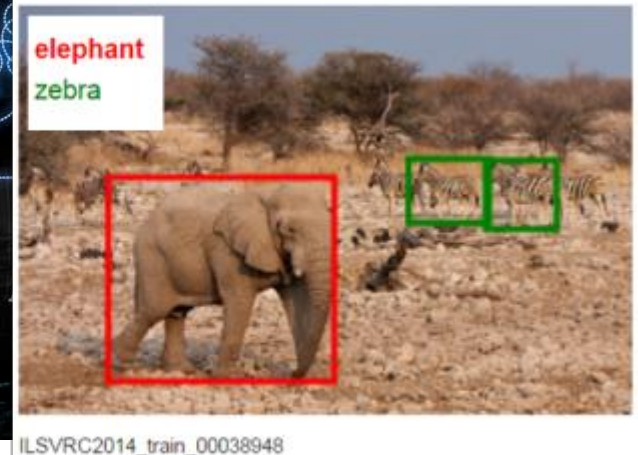
15

□ Applications can't program by hand

- ▣ Driverless car
- ▣ Handwriting recognition
- ▣ Natural language processing
- ▣ Computer vision



□ Understanding





What is machine learning useful for?

Automatic speech recognition

自动语音识别

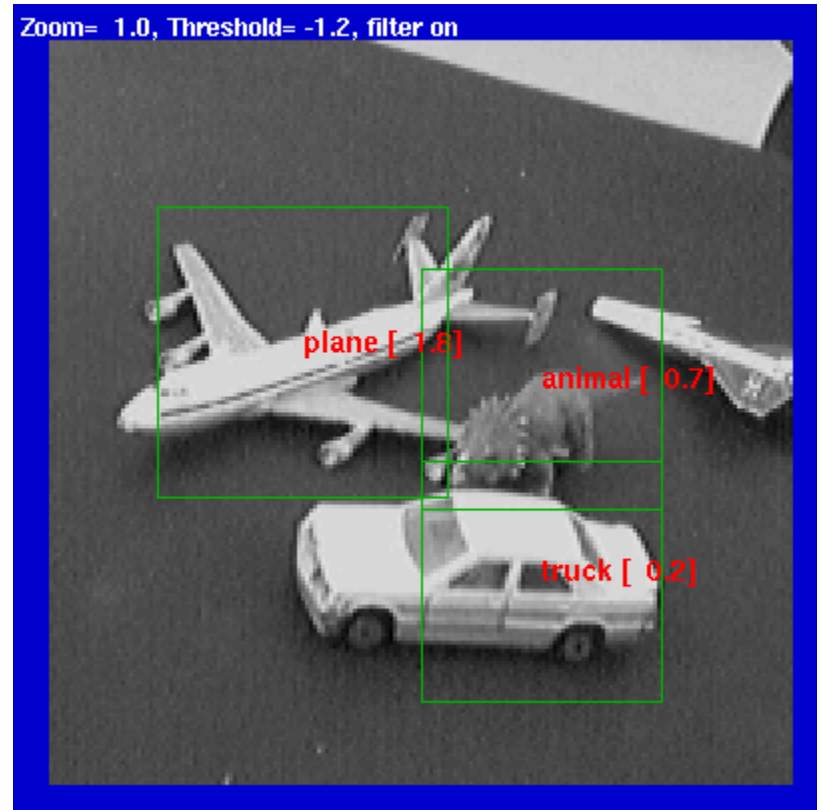
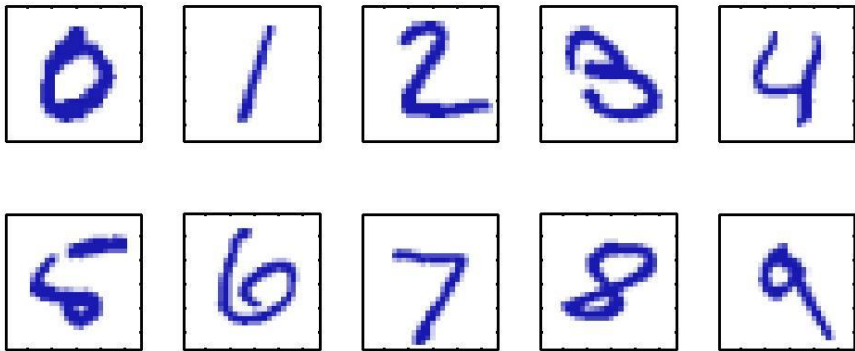
17

Now most **Speech Recognizers or Translators** are able to learn — the more you play/use them, the smarter they become



Computer vision: e.g. object, face and handwriting recognition

18



19

Relations between pages

06/10/04 15:44

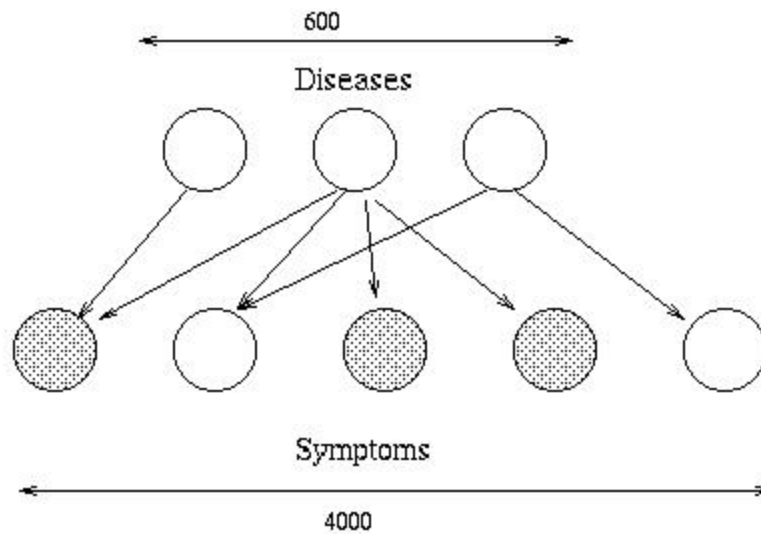
Financial prediction

20



Medical diagnosis (医学诊断)

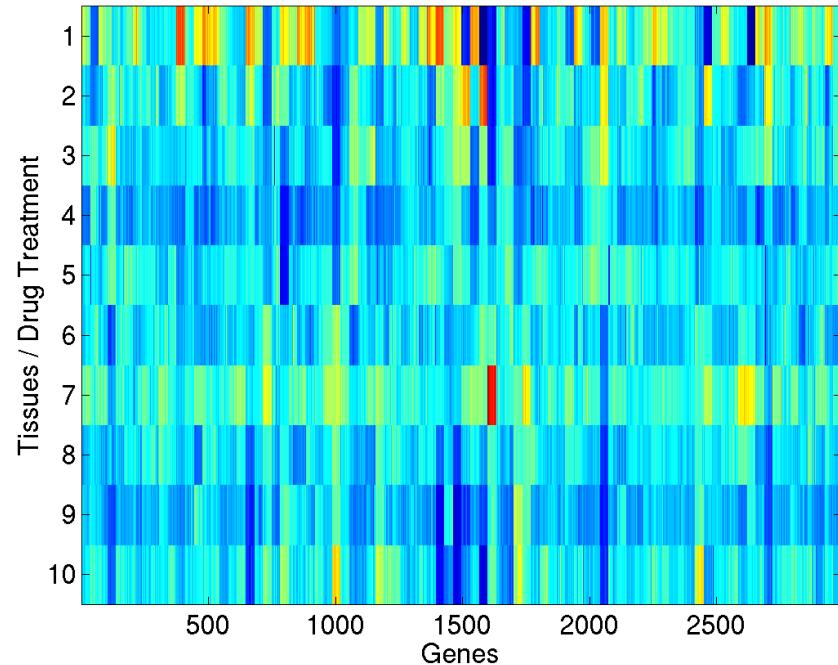
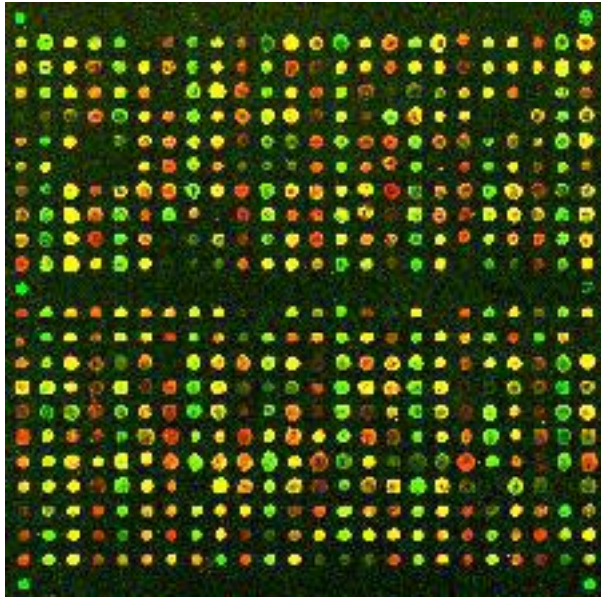
21



(image from Kevin Murphy)

Bioinformatics (生物信息学)

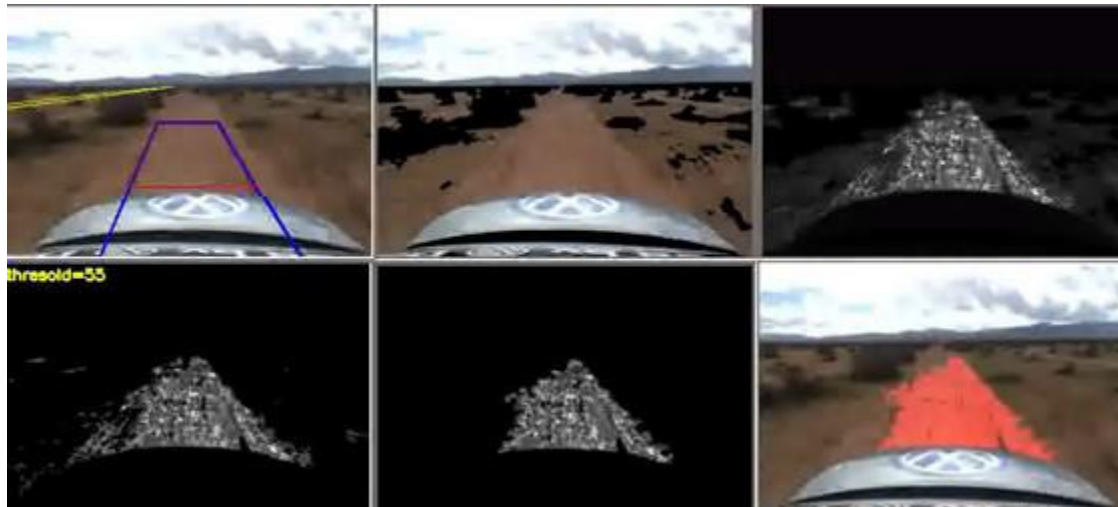
22



e.g. modeling gene microarray (微阵列) data, protein structure prediction

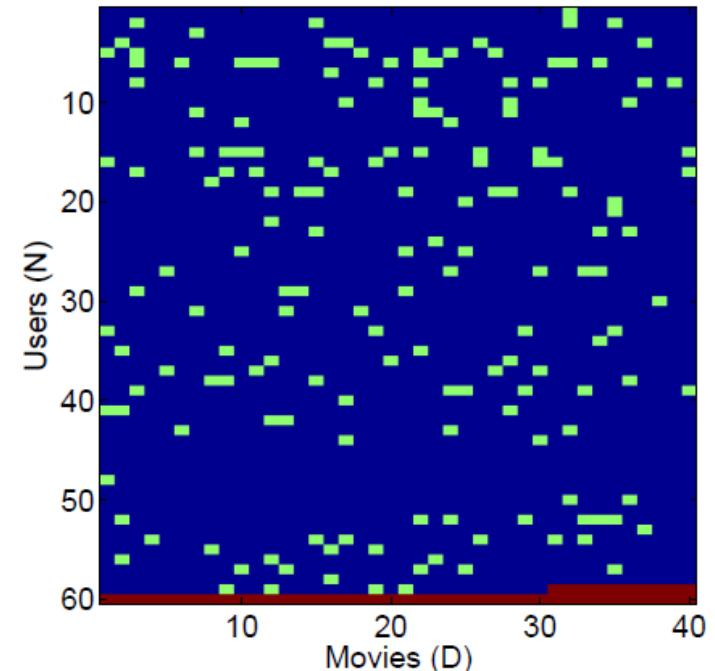
Robotics

23



Movie recommendation systems

24



Challenge: to improve the accuracy of movie preference predictions
Netflix \$1m Prize.

Machine Learning

25

***Machine learning** is an interdisciplinary field focusing on both the mathematical foundations and practical applications of systems that learn, reason and act.*

*机器学习是一个交叉学科的领域，着重于研究具有学习、推理和行动的
系统所需要的数学基础以及实际应用*

Other related terms: Pattern Recognition（模式识别），Neural Networks（神经网络），Data Mining（数据挖掘），Statistical Modeling（统计模型） ...

Using ideas from: Statistics, Computer Science, Engineering, Applied Mathematics, Cognitive Science（认知科学），Psychology（心理学），Computational Neuroscience（计算神经学），Economics

The goal of these lectures: to introduce important concepts, models and algorithms in machine learning.

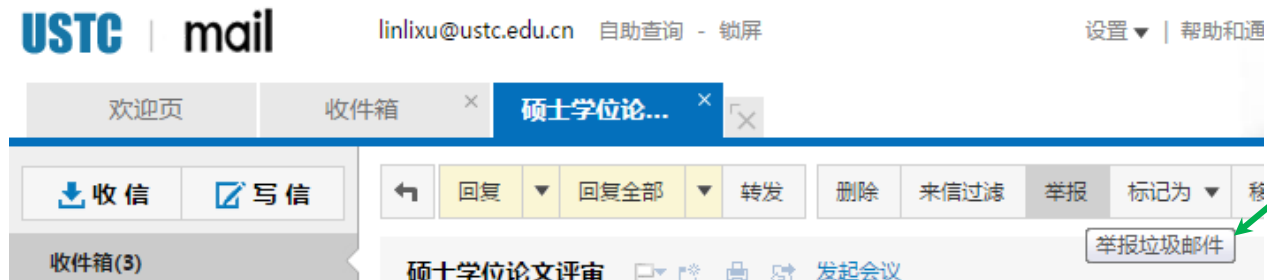
Machine Learning: Definition

26

- Tom Mitchell (1998) Well-posed Learning Problem: A computer program is said to *learn* from **experience E** with respect to some **task T** and some **performance measure P**, if its performance on **T**, as measured by **P**, improves with **experience E**.

“A computer program is said to *learn* from **experience E** with respect to some **task T** and some **performance measure P**, if its performance on T, as measured by P, improves with experience E.”

Suppose your email program watches which emails you do or do not mark as spam, and based on that learns how to better filter spam. What is the task T in this setting?



- ❑ ☒ Classifying emails as spam or not spam. **T**
- ❑ Watching you label emails as spam or not spam. **E**
- ❑ The number (or fraction) of emails correctly classified as spam/not spam. **P**
- ❑ None of the above—this is not a machine learning problem.

Types of Learning

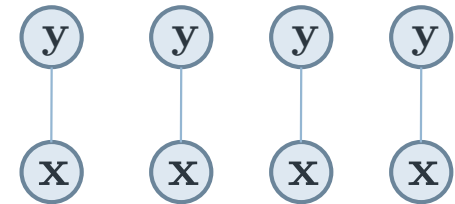
28

Imagine an agent or machine which experiences a series of sensory inputs:

$$x_1, x_2, x_3, x_4, \dots$$

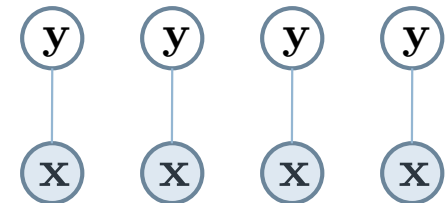
Supervised learning (监督学习) :

The machine is also given desired outputs y_1, y_2, \dots , and its goal is to learn to produce the correct output given a new input.

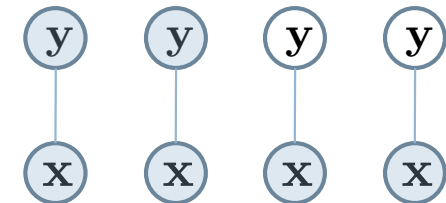


Unsupervised learning (无监督学习) :

outputs y_1, y_2, \dots Not given, the agent still wants to build a model of x that can be used for reasoning, decision making, predicting things, communicating etc.



Semi-supervised learning (半监督学习)



Representing “objects” in machine learning

29

- An **example** or **instance**, x , represents a specific object
- x often represented by a d -dimensional feature vector $x = (x_1, \dots, x_d) \in \mathbb{R}^d$
- Each dimension is called a feature or attribute
- Continuous or discrete
- x is a point in the d -dimensional feature space
- Abstraction of object. Ignores any other aspects (e.g., two people having the same weight and height may be considered identical)

Feature vector representation

30

- Text document *one-hot 向量.*
 - Vocabulary of size d ($\sim 100,000$)
 - “bag of words”: counts of each vocabulary entry *词袋*
 - Often remove stopwords: the, of, at, in, ...
 - Special “out-of-vocabulary” (OOV) entry catches all unknown words

缺点: 未表示出单词 词的相关性

Feature vector representation

31

□ Text document



word \longrightarrow vector

word 2 vec

Analogy: Beijing-China=Paris-France

类比

Feature vector representation

32

- Image
 - ▣ Pixels, Color histogram
- Feature extraction using convolution

1 _{x1}	1 _{x0}	1 _{x1}	0	0
0 _{x0}	1 _{x1}	1 _{x0}	1	0
0 _{x1}	0 _{x0}	1 _{x1}	1	1
0	0	1	1	0
0	1	1	0	0

Image

4		

Convolved
Feature

Feature vector representation

33

- Software
 - ▣ Execution profile: the number of times each line is executed
- Bank account
 - ▣ Credit rating, balance, #deposits in last day, week, month, year, #withdrawals, ...
- You and me
 - ▣ Medical test1, test2, test3, ...

Key Ingredients

34

Data

The data set D consists of N data points:

$$D = \{x_1, x_2, \dots, x_N\}$$

Predictions (预测)

We are generally interested in predicting something based on the observed data set.

Given D what can we say about x_{N+1} ?

Model

To make predictions, we need to make some assumptions. We can often express these assumptions in the form of a model, with some parameters (参数)

这些假设很重要 (基于数据观察而认为较合理的条件), 可用于建模
(类似于朴素贝叶斯中条件独立性假设)

Given data D , we learn the model parameters, from which we can predict new data points.

Key Ingredients

35

$$\min_f \text{Loss}(Y, f(X))$$

模型

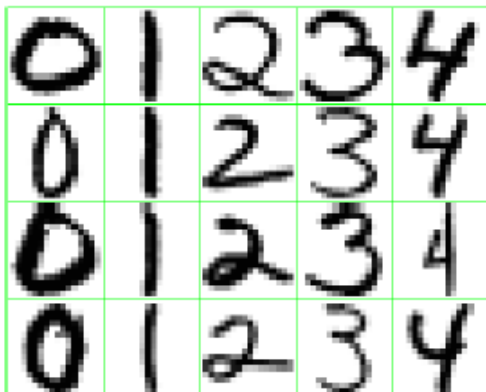
→ $f(x)$

预测: $y_{\text{new}} = f(\text{3})$

输入 X

输出 Y

数据 →

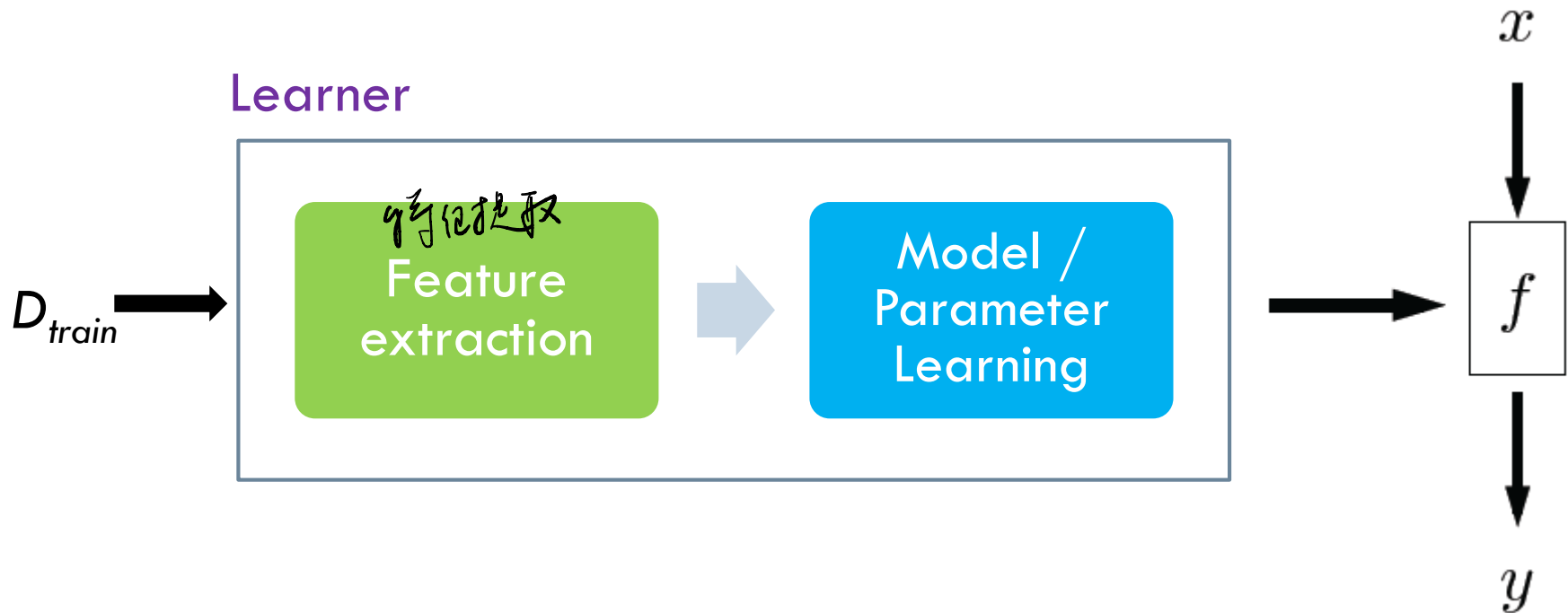


digits recognition;

$$\mathcal{Y} = \{0, \dots, 9\}$$

Learning Framework

36



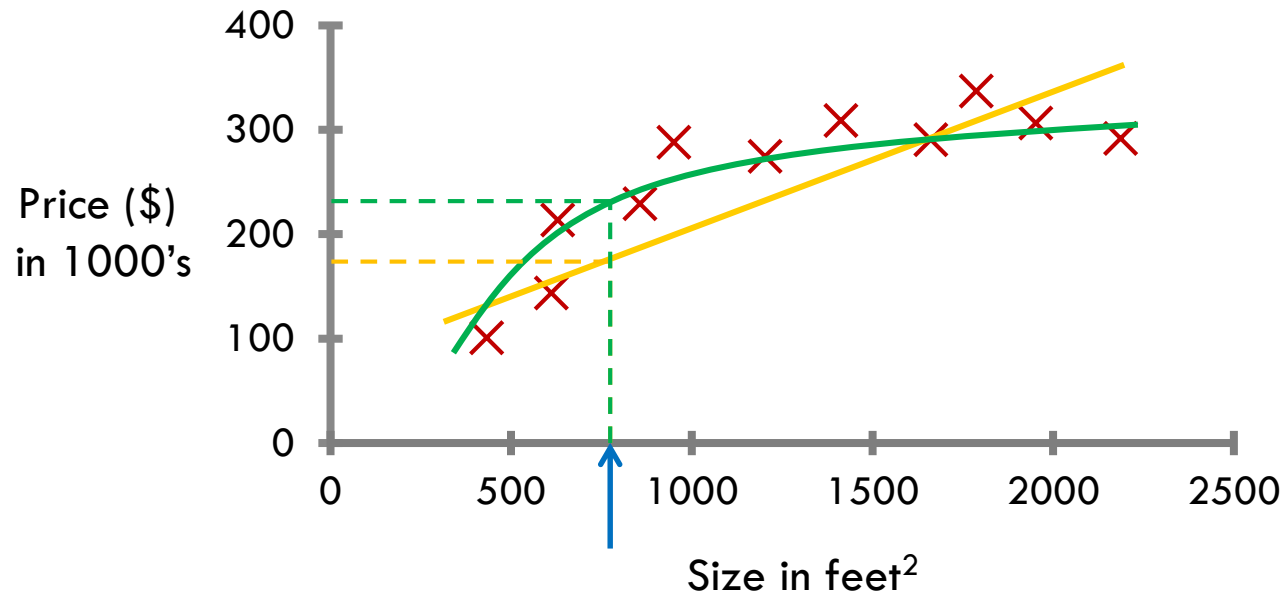
37

Learning Problems

Housing price prediction

38

房价预测



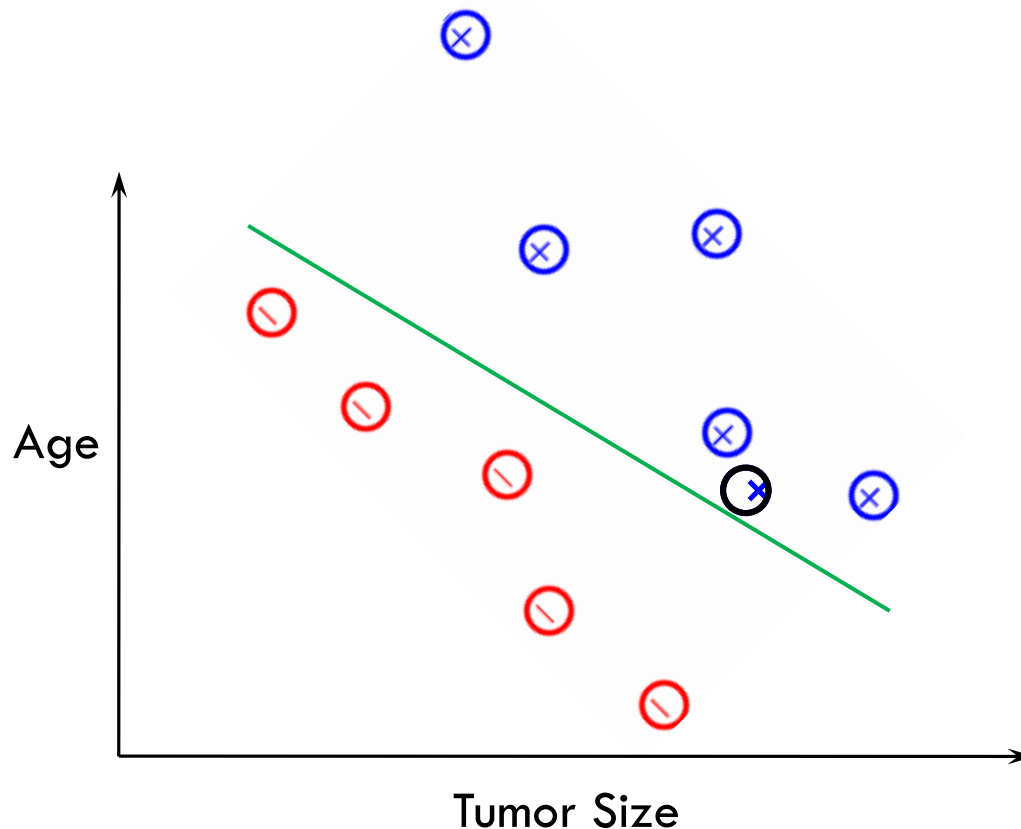
Supervised Learning

“right answers” given

Regression (回归) : Predict
continuous valued output (price)

Breast cancer (malignant, benign)

39



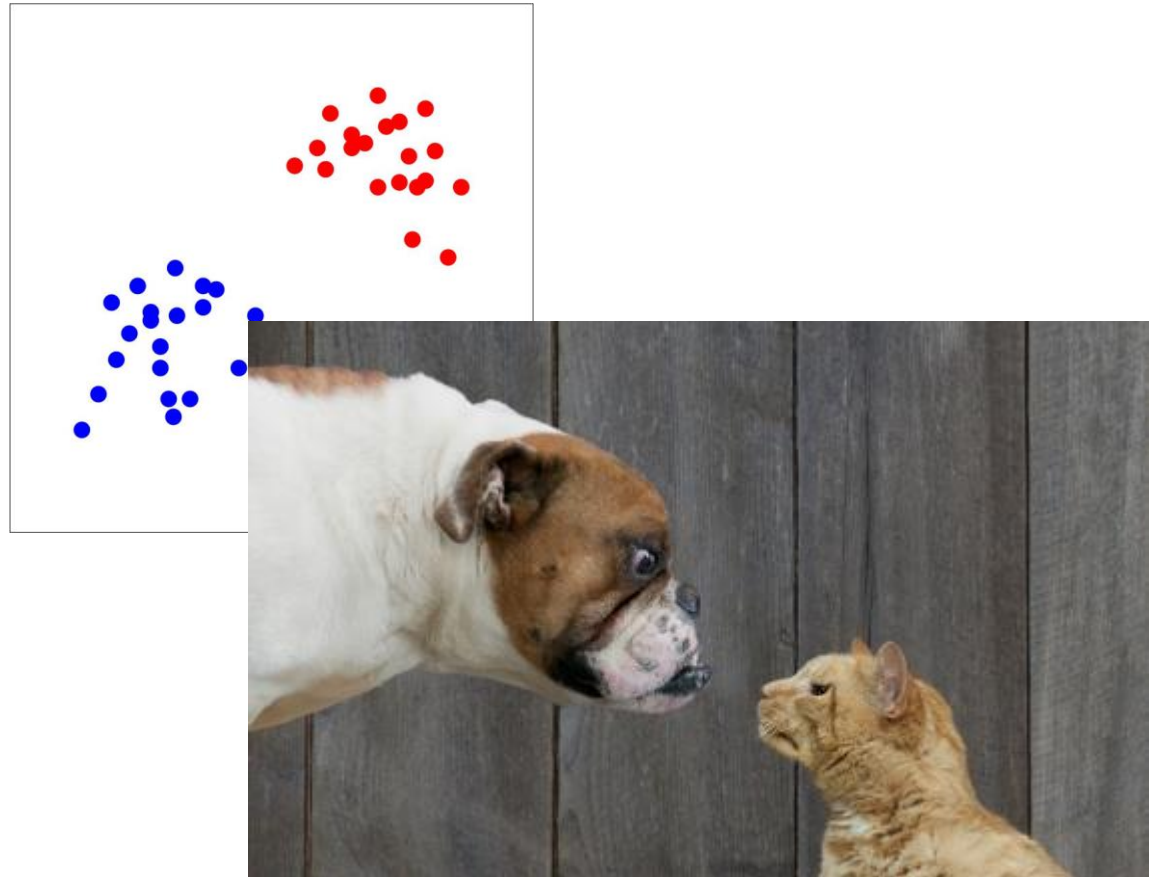
Supervised Learning

“right answers” given

Classification (分类) : Predict
discrete valued output

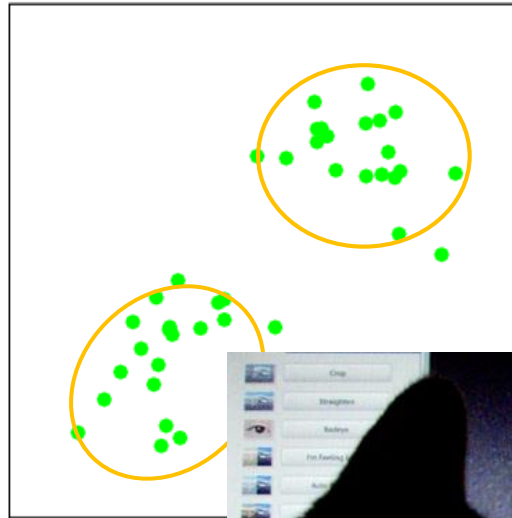
Supervised Learning

40



Unsupervised Learning

41



Next...

42

- Machine learning algorithms
 - ▣ Supervised learning
 - ▣ Unsupervised learning

