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

- 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

Learning is essential for unknown environments,

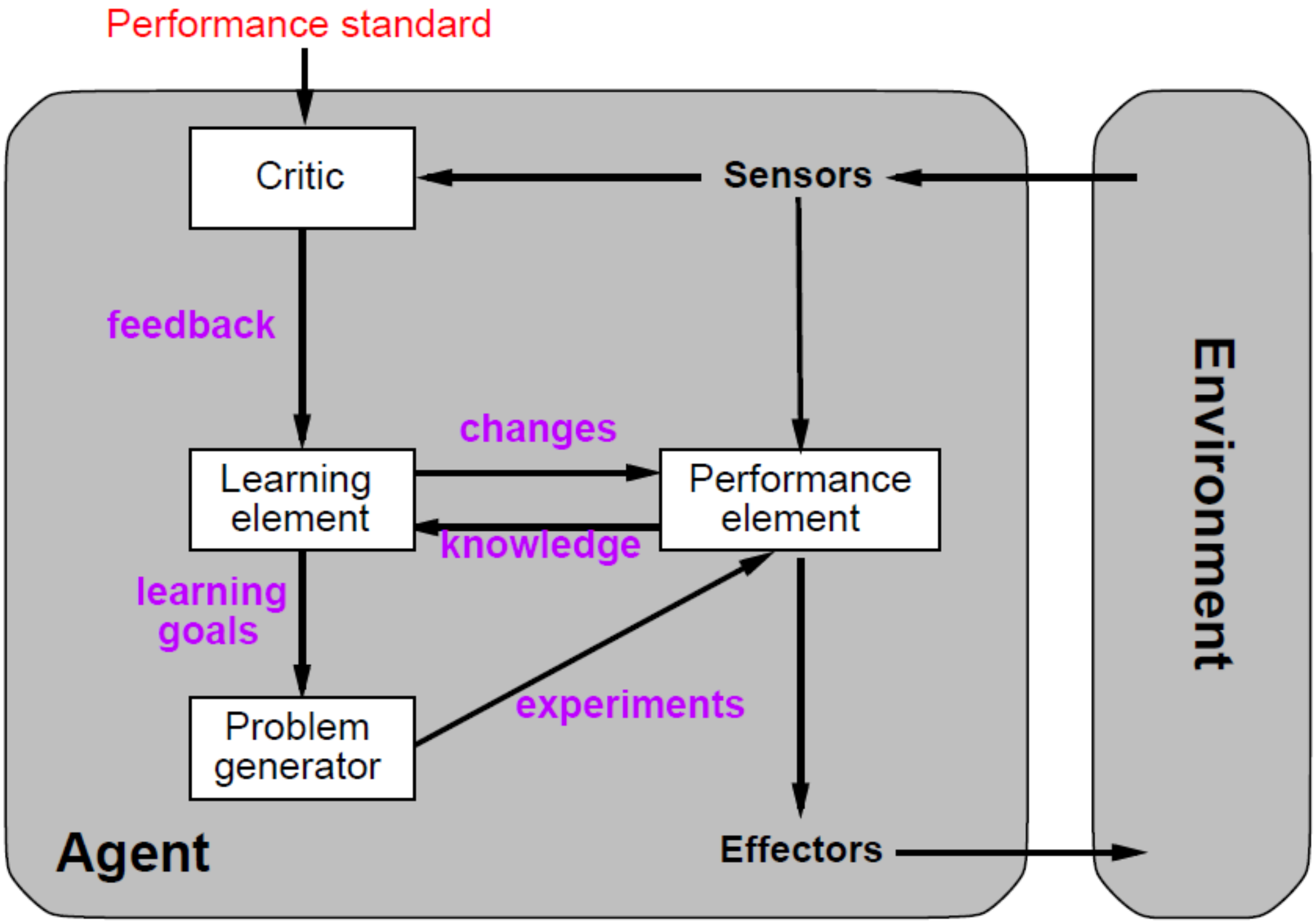
- i.e., when designer lacks omniscience（全知）

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



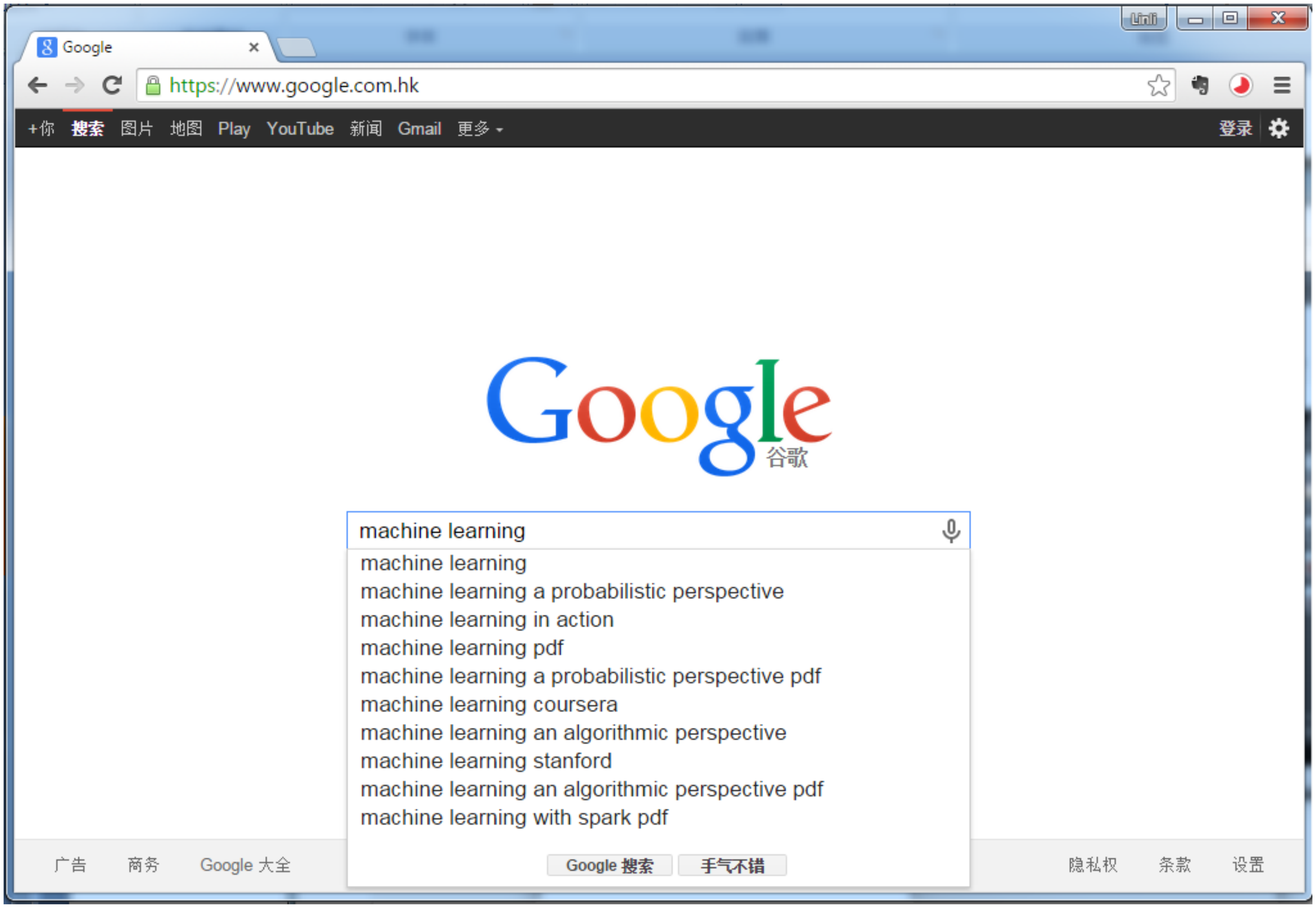
Learning element

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

Introduction to Machine Learning

Machine Learning Everyday: Search Engine



Machine Learning Everyday: Spam Detection （垃圾邮件检测）

Machine Learning Everyday: Machine Translation

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 مستعدّة غير أنّ اللحمة ضعيفة.	The alcohol is ready nevertheless the meat is weak.

Statistical machine translation models

Machine Learning Everyday: Face Detection

- Now in most digital cameras for auto focusing

Also blink and smile detection!

Machine Learning

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

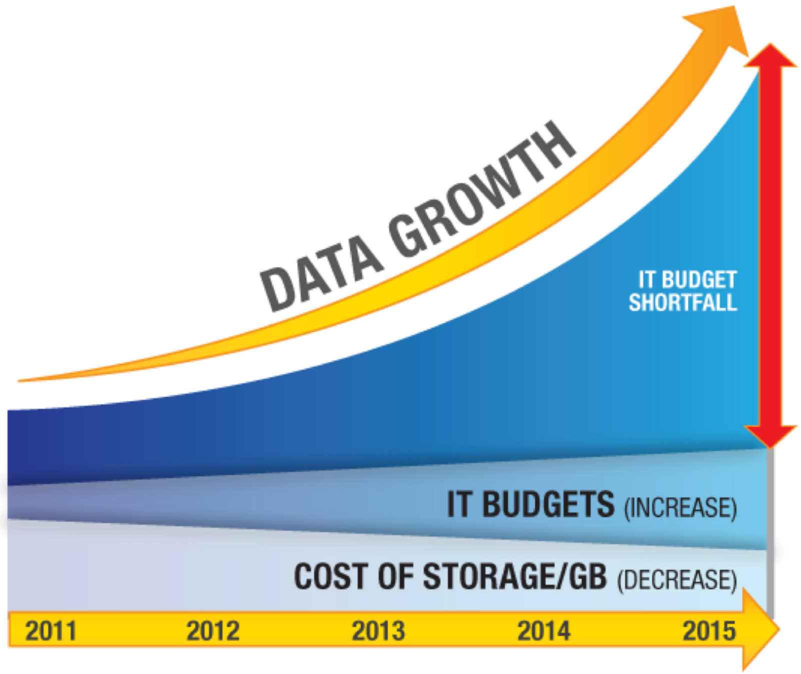
Why Machine Learning?

- 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”)

…

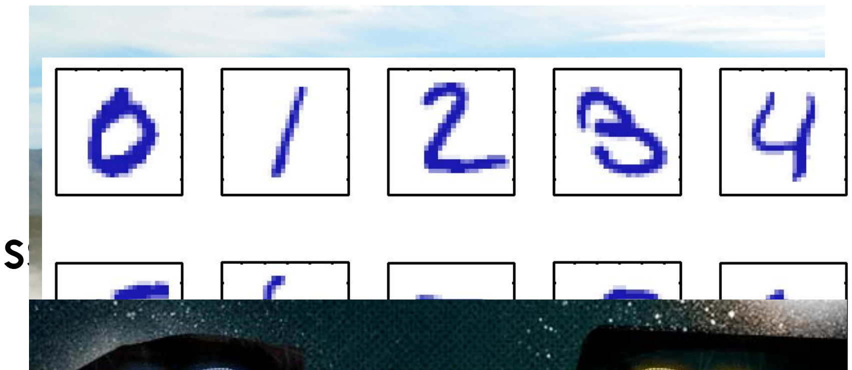
Why Machine Learning?

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

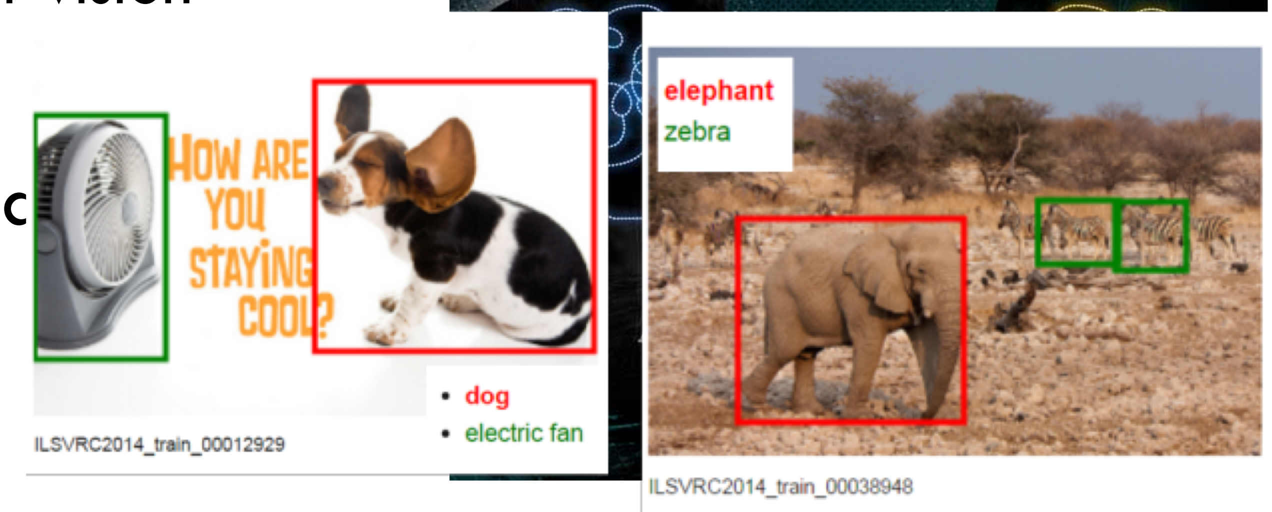


Why Machine Learning?

- Applications can't program by hand
  - Driverless car
  - Handwriting recognition
  - Natural language proces
  - Computer vision



- Understanc

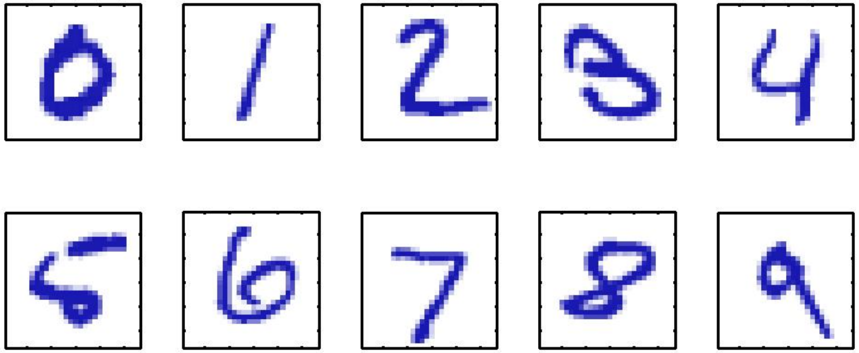


What is machine learning useful for?

Automatic speech recognition
自动语音识别

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



Information retrieval—信息检索

Reading, digesting, and categorizing a vast text database is too much for human

Web Pages

Retrieval（检索）

Categorization（分类）

Clustering（聚类）

Relations between pages


Google Search: Unsupervised Learning http://www.google.com/search?q=Unsupervised+Learning&sourceid=fir...

Web Images Groups News Froogle more »

Unsupervised Learning Search Advanced Search Preferences

Web Results 1 - 10 of about 150,000 for Unsupervised Learning. (0.27 seconds)

Mixture modelling, Clustering, Intrinsic classification ...
Mixture Modelling page. Welcome to David Dowe's clustering, mixture modelling and unsupervised learning page. Mixture modelling (or ...
www.csse.monash.edu.au/~dld/mixture.modelling.page.html - 26k - 4 Oct 2004 - Cached - Similar pages

ACL'99 Workshop -- Unsupervised Learning in Natural Language ...
PROGRAM. ACL'99 Workshop Unsupervised Learning in Natural Language Processing. University of Maryland June 21, 1999. Endorsed by SIGNLL ...
www.ai.sri.com/~kehler/unsup-acl-99.html - 5k - Cached - Similar pages

Unsupervised learning and Clustering
cgm.cs.mcgill.ca/~soss/cs644/projects/wijhe/ - 1k - Cached - Similar pages

NIPS*98 Workshop - Integrating Supervised and Unsupervised ...
NIPS*98 Workshop "Integrating Supervised and Unsupervised Learning" Friday, December 4, 1998. ... 4:45-5:30, Theories of Unsupervised Learning and Missing Values. ...
www-2.cs.cmu.edu/~mccallum/supunsup/ - 7k - Cached - Similar pages

NIPS Tutorial 1999
Probabilistic Models for Unsupervised Learning Tutorial presented at the 1999 NIPS Conference by Zoubin Ghahramani and Sam Roweis. ...
www.gatsby.ucl.ac.uk/~zoubin/NIPStutorial.html - 4k - Cached - Similar pages

Gatsby Course: Unsupervised Learning : Homepage
Unsupervised Learning (Fall 2000). ... Syllabus (resources page): 10/10 1 - Introduction to Unsupervised Learning Geoff project: (ps, pdf). ...
www.gatsby.ucl.ac.uk/~quaid/course/ - 15k - Cached - Similar pages
[More results from www.gatsby.ucl.ac.uk]

[PDF] Unsupervised Learning of the Morphology of a Natural Language
File Format: PDF/Adobe Acrobat - View as HTML
Page 1. Page 2. Page 3. Page 4. Page 5. Page 6. Page 7. Page 8. Page 9. Page 10. Page 11. Page 12. Page 13. Page 14. Page 15. Page 16. Page 17. Page 18. Page 19 ...
acl.ldc.upenn.edu/J/J01/J01-2001.pdf - Similar pages

Unsupervised Learning - The MIT Press
... From Bradford Books: Unsupervised Learning Foundations of Neural Computation Edited by Geoffrey Hinton and Terrence J. Sejnowski Since its founding in 1989 by ...
mitpress.mit.edu/book-home.tcl?isbn=026258168X - 13k - Cached - Similar pages

[PS] Unsupervised Learning of Disambiguation Rules for Part of
File Format: Adobe PostScript - View as Text
Unsupervised Learning of Disambiguation Rules for Part of. Speech Tagging. Eric Brill. 1. ... It is possible to use unsupervised learning to train stochastic. ...
www.cs.jhu.edu/~brill/acl-wksh.ps - Similar pages

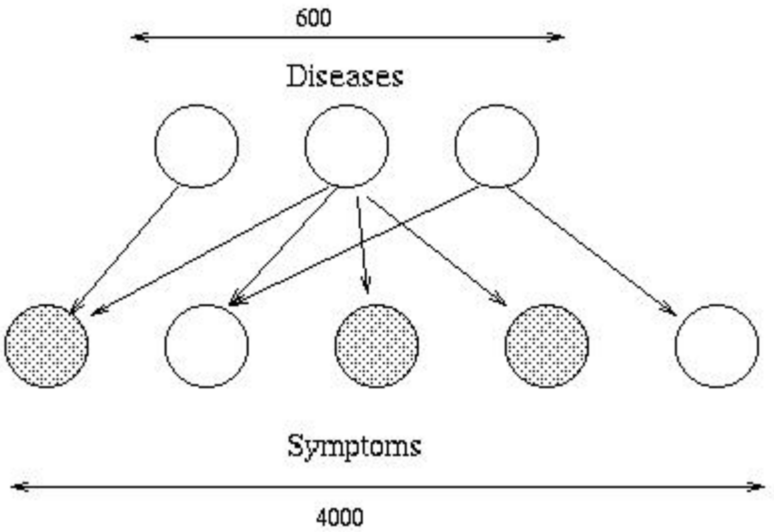
The Unsupervised Learning Group (ULG) at UT Austin
The Unsupervised Learning Group (ULG). What ? The Unsupervised Learning Group (ULG) is a group of graduate students from the Computer ...
www.lans.ece.utexas.edu/ulg/ - 14k - Cached - Similar pages

Gooooooooooogle ▸
Result Page: 1 2 3 4 5 6 7 8 9 10 Next

1 of 2 06/10/04 15:44

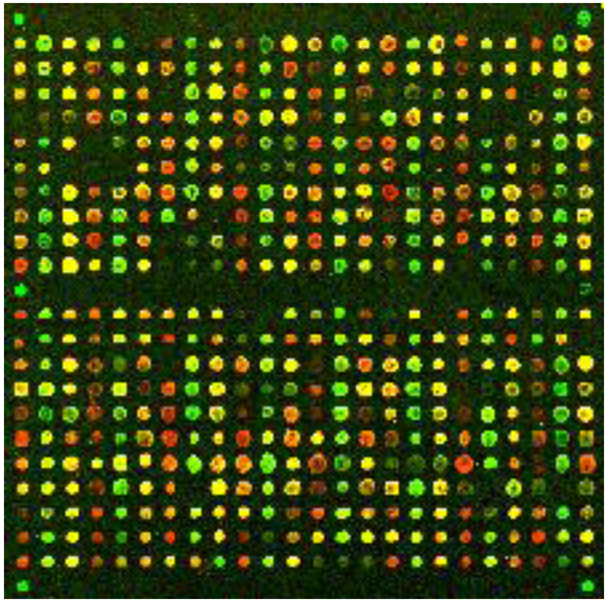
Financial prediction

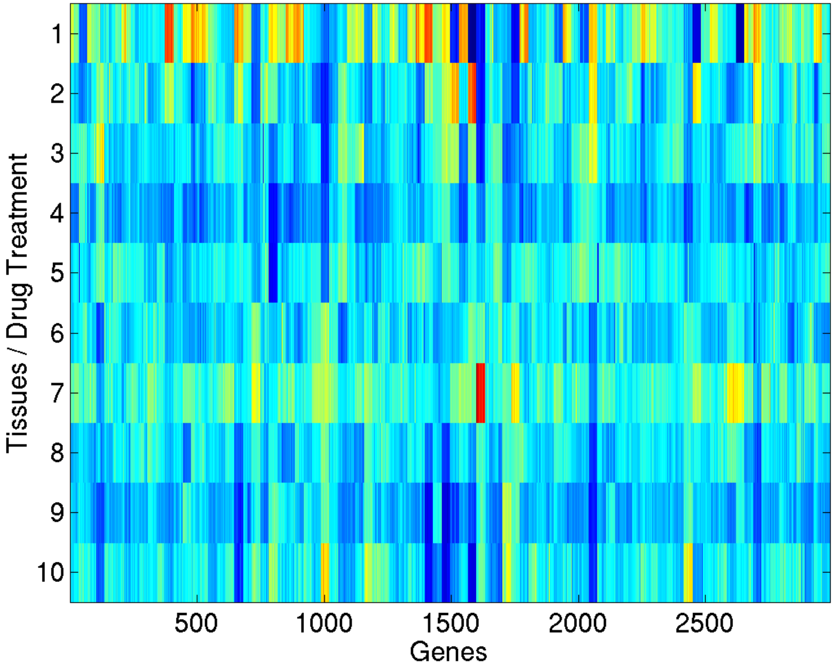
Medical diagnosis（医学诊断）



(image from Kevin Murphy)

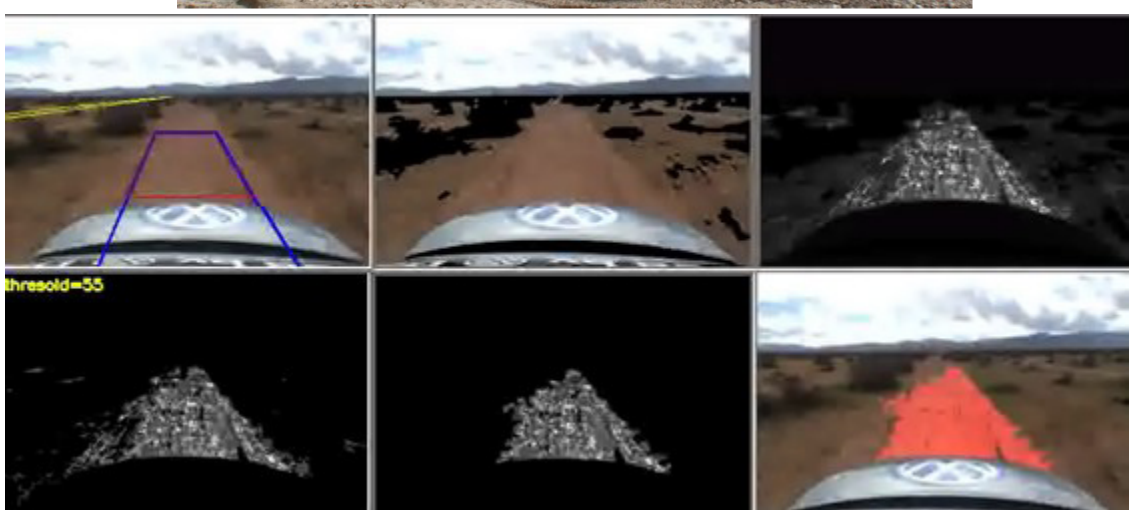
Bioinformatics（生物信息学）



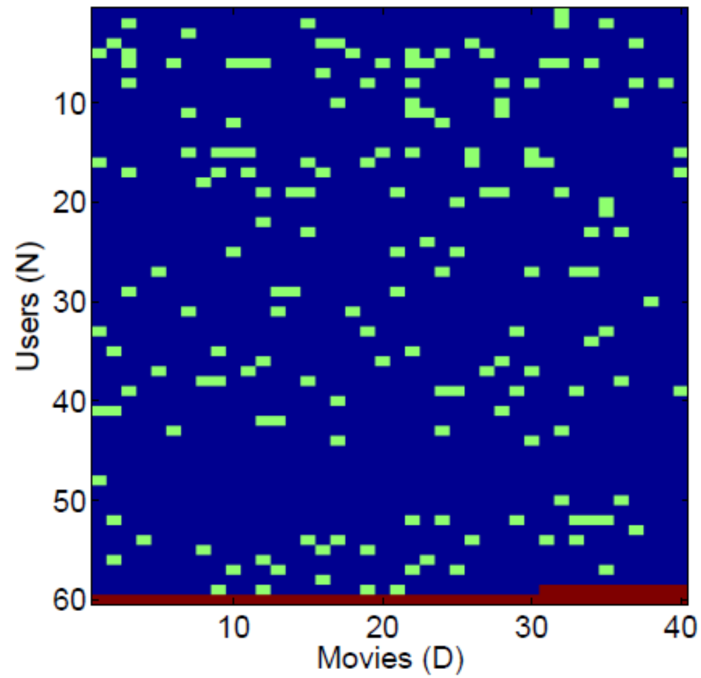


e.g. modeling gene microarray（微阵列）data, protein structure prediction

Robotics



Movie recommendation systems



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

Machine Learning

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

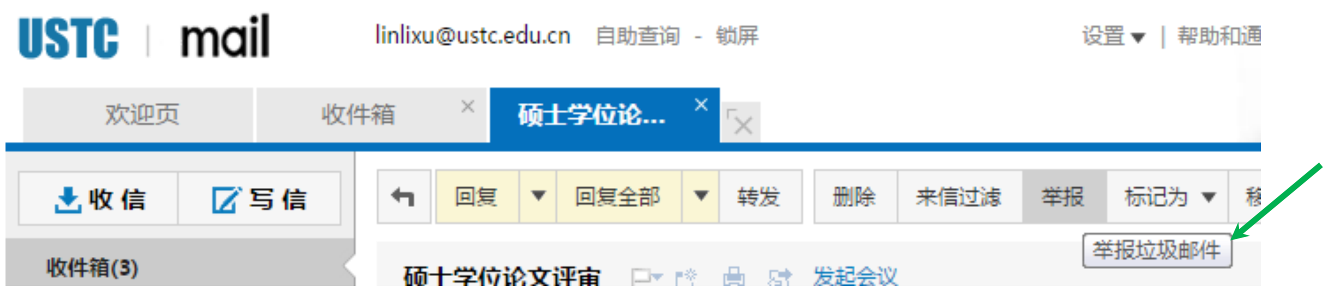
The goal of these lectures: to introduce important concepts, models and algorithms in machine learning.

Machine Learning: Definition

- 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.

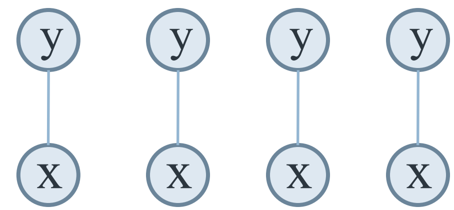
Types of Learning

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

$x_1, x_2, x_3, x_4, \ldots$

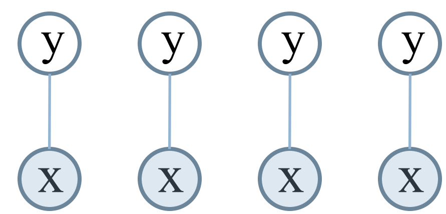
Supervised learning（监督学习）:

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

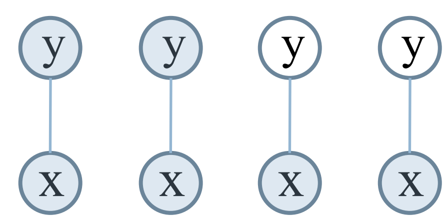


Unsupervised learning（无监督学习）:

outputs $y_1, y_2, \ldots$ 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

- An example or instance, x, represents a specific object
- x often represented by a d-dimensional feature vector $x = (x_1, \ldots, x_d) \in R^d$
- 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

- Text document
  - Vocabulary of size d (~100,000)
  - "bag of words": counts of each vocabulary entry
  - Often remove stopwords: the, of, at, in, …
  - Special "out-of-vocabulary" (OOV) entry catches all unknown words

Feature vector representation

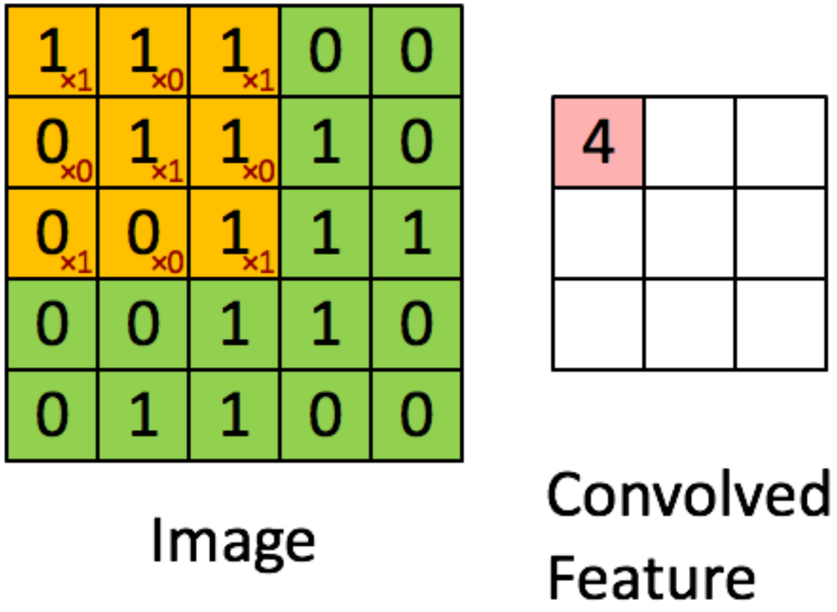
- Text document

word → vector

Analogy: Beijing-China=Paris-France

Feature vector representation

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



Feature vector representation

- 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

Data

The data set D consists of N data points:

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

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）

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

Key Ingredients

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

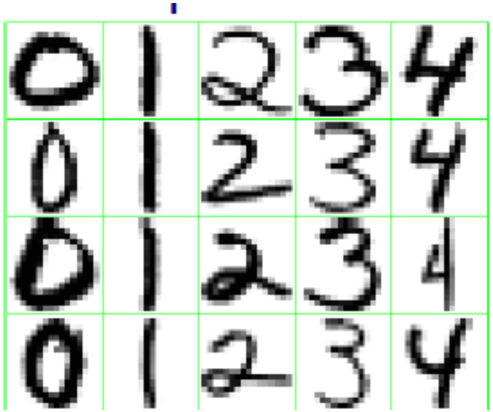
模型 → f(x)

预测： $y_{new}=f($  $)$

输入X

输出Y

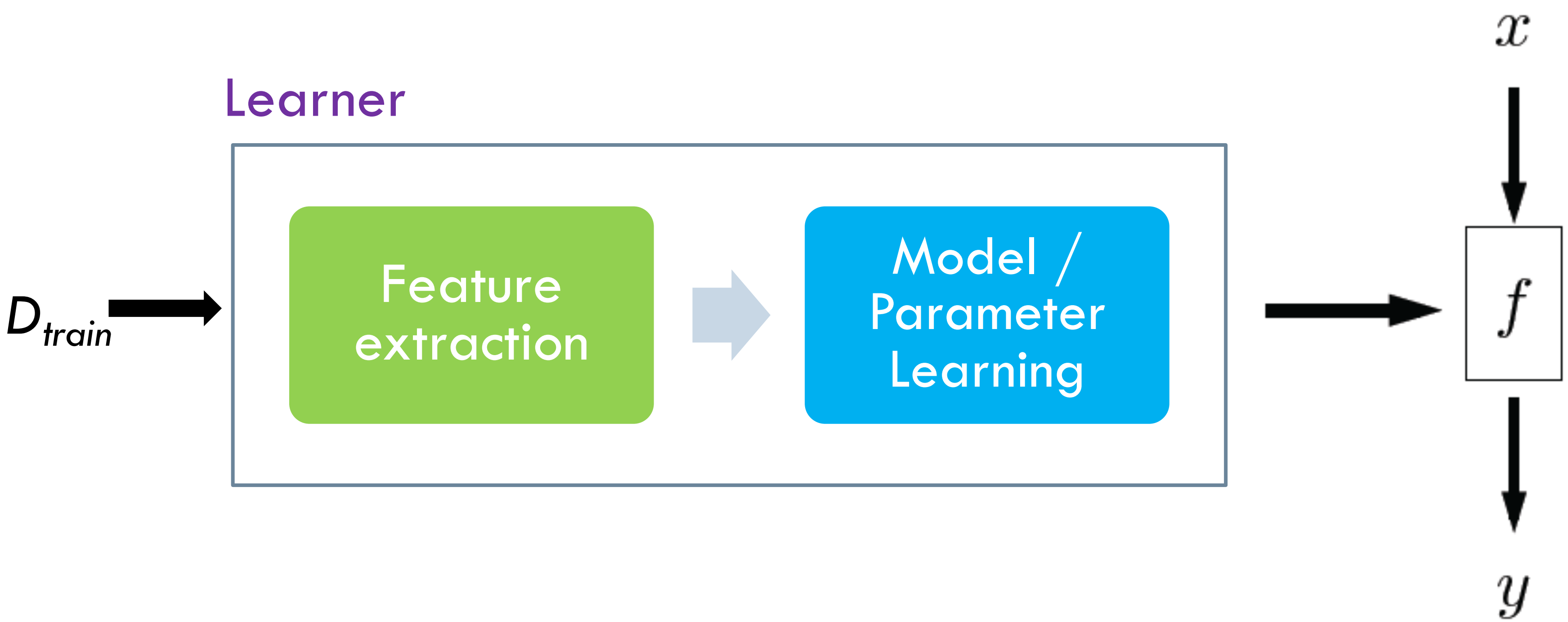
数据 →



digits recognition;

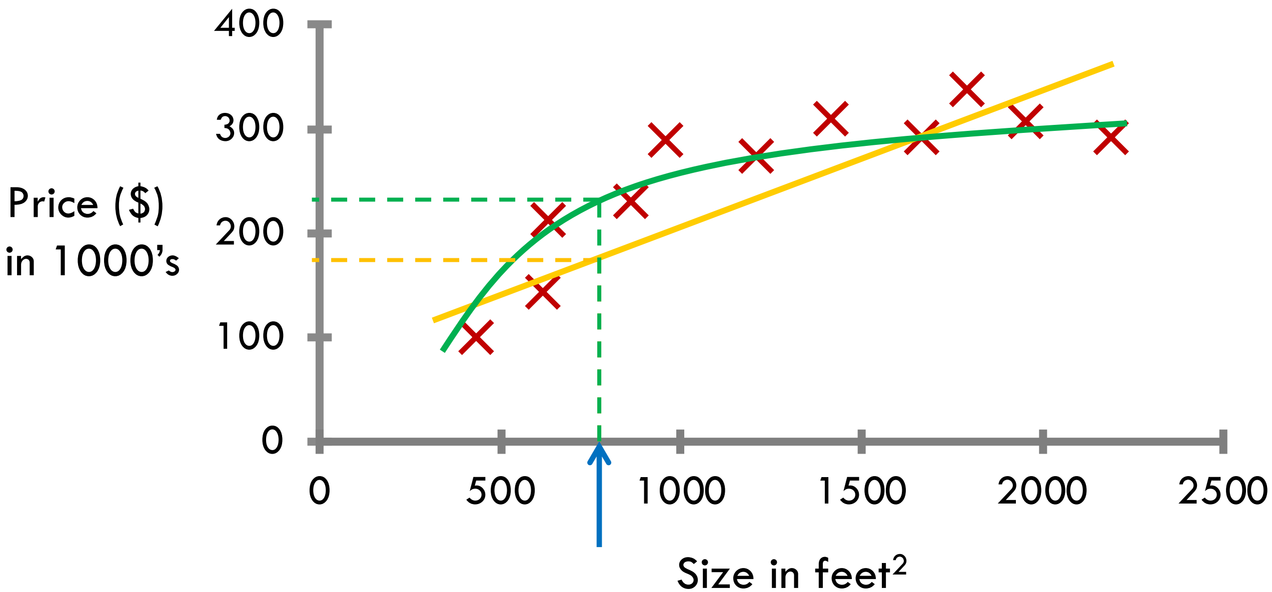
$\mathcal{Y} = \{0, \dots, 9\}$

Learning Framework



Learning Problems

Housing price prediction

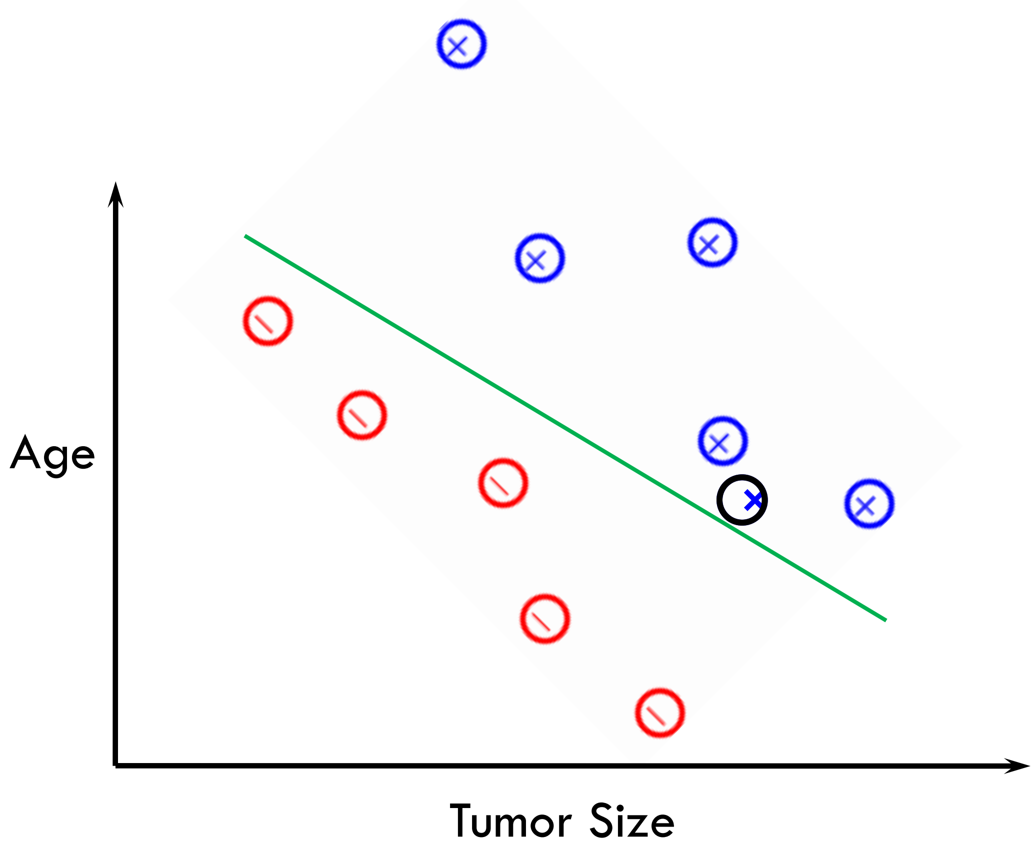


Supervised Learning

"right answers" given

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

Breast cancer (malignant, benign)

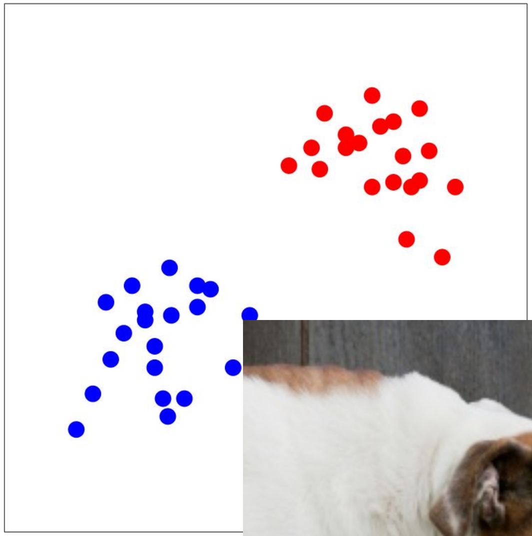


Supervised Learning

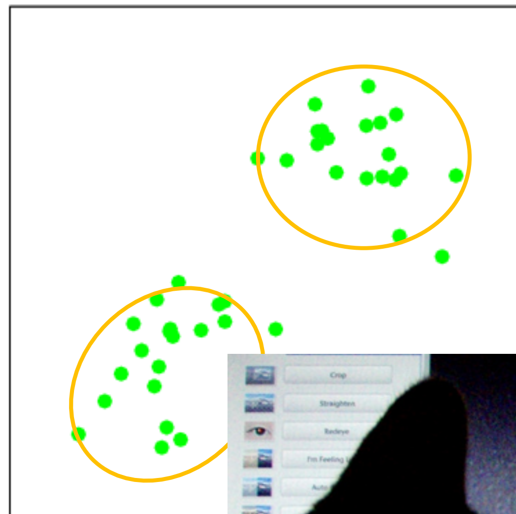
"right answers" given

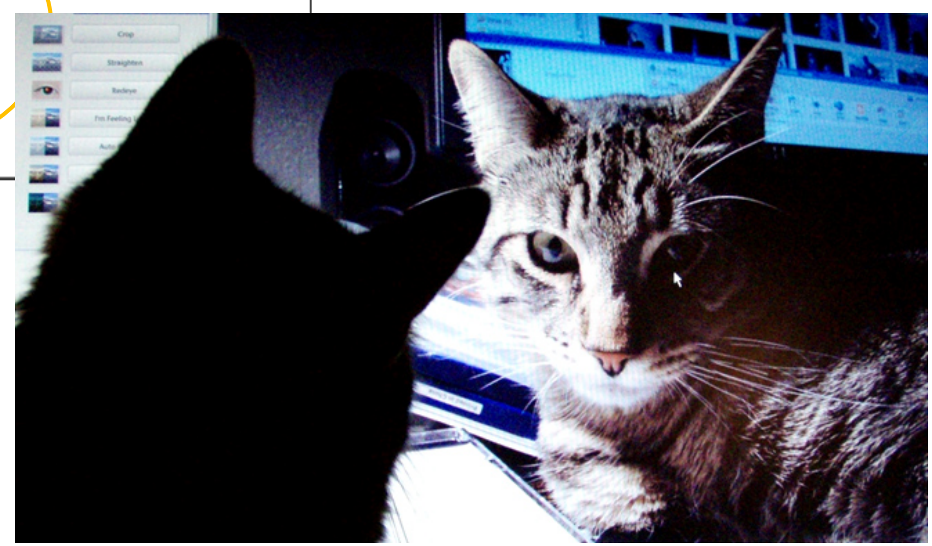
Classification（分类）：Predict discrete valued output

Supervised Learning



Unsupervised Learning





Next…

- Machine learning algorithms
  - Supervised learning
  - Unsupervised learning