



# 统计机器学习 (小班研讨)

主讲教师：刘峤

# 课程说明

- 课程编号：0908357013
- 课程性质：专业选修课      学分：**2**
- 总学时：**40** （课堂讲授：课堂研讨  $\approx$  1:1）
- 上课时间地点：
  - 2019年春季 **1 - 8**周
  - 周一**第9 ~ 11节**，主楼中（**508**）
  - 周三**第3 ~ 4节**，主楼中（**508**）

# 课程说明

- **成绩构成：**

- **研讨成绩：20%**（积极主动提问和参与讨论）
- 平时作业：**20%**（独立完成2次作业） - 我们用Python3
- **综合设计：30%**（完成一个课程项目）
- 期末考试：**30%**（考试：闭卷、笔试）

- **联系方式：**

- 邮箱：**qliu@uestc.edu.cn**



# Course goals

---

## ◆ Be familiar with major ML methods

- Regression, Decision trees & random forests
- Naive Bayes, HMMs, SVM, kernels, PCA, LDA
- Deep learning ...

## ◆ Know their strengths and weaknesses

- know jargon, concepts, theory
- be able to modify and code algorithms

## ◆ Know how to do research -- preliminarily

- be able to read literature and write research proposal



# Dynamic Course Plan



**Please Don't ... not really helpful 😊**



# Introductions

---

◆ If you're **waiting** to get into this course

- **It won't happen as per your good wishes ... 😊**
- But the course will be offered again in the next spring

◆ Alternate courses

- 李晓渝：大数据分析
- 蓝 天：机器学习

# We will use Python3

---

## ◆ is Python a better language than R?

- Refer to : **Data Science Wars: R vs Python**

- <https://www.datacamp.com/community/tutorials/r-or-python-for-data-analysis>

- R has a much bigger library of statistical packages
- Python is better for building analytics tools
- Python is better for deep learning
- Python relies on a few main packages, R has hundreds
- Python has **Jupyter Notebook** !



# 教材与参考书推荐

- 课程教材

- 《Pattern Recognition and Machine Learning》 2007

- Christopher M. Bishop

- 参考书推荐

- 《The Elements of Statistical Learning: Data Mining, Inference, and Prediction》 2009

- Trevor Hastie, Robert Tibshirani, Jerome Friedman

- 《Learning From Data》 2012

- Yaser S. Abu-Mostafa, Malik Magdon-Ismael, Hsuan-Tien Lin



# 公开课推荐

---

## ◆ Machine Learning - Hsuan-Tien Lin, 林軒田

- Machine Learning Foundations
- Machine Learning Techniques
- <https://www.csie.ntu.edu.tw/~htlin/mooc/>

## ◆ Machine Learning - Andrew Ng

- <https://zh.coursera.org/learn/machine-learning>

## ◆ Machine Learning - Geoffrey Hinton

- <https://zh.coursera.org/learn/machine-learning>



# Useful Links and Resources

---

## ◆ Python

- **Nick Parlante: Google's Python Class**

- <https://developers.google.com/edu/python/>

## ◆ Linear Algebra Resources

- **Linear Algebra lectures -- Gil Strang @ MIT**

- **A Tutorial on Linear Algebra -- C. T. Abdallah**

- <http://www.seas.upenn.edu/~jadbabai/ESE504/LAreview.pdf>

- **The Matrix Cookbook**

- <http://matrixcookbook.com/>



# Useful Links and Resources

---

## ◆ Probability Resources

- **Review of probability -- David Blei**

- [http://www.cs.princeton.edu/courses/archive/spring07/cos424/scribe\\_notes/0208.pdf](http://www.cs.princeton.edu/courses/archive/spring07/cos424/scribe_notes/0208.pdf)

- **Andrew Moore's Probability Tutorial**

- <http://www.autonlab.org/tutorials/prob.html>

- **Another probability review, from UCI**

- <http://www.ics.uci.edu/~smyth/courses/cs274/notes/notes1.pdf>



# 第1章 绪论

## An Overview of Machine Learning

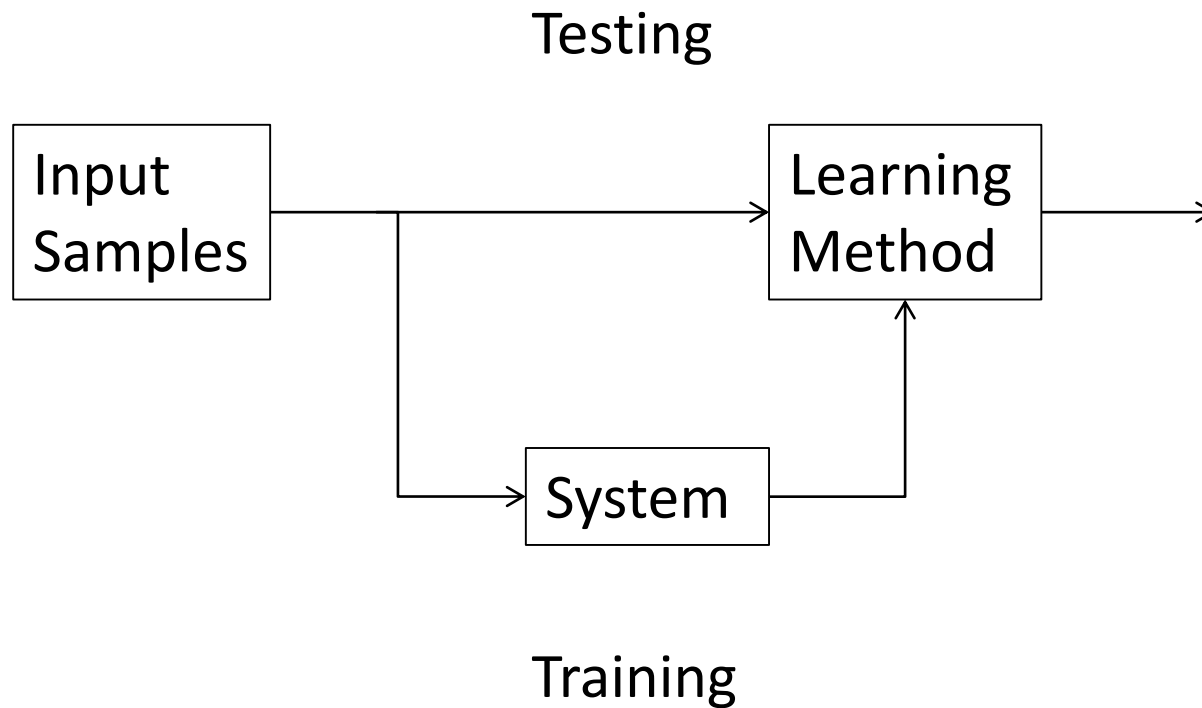
# What is Machine Learning?

---

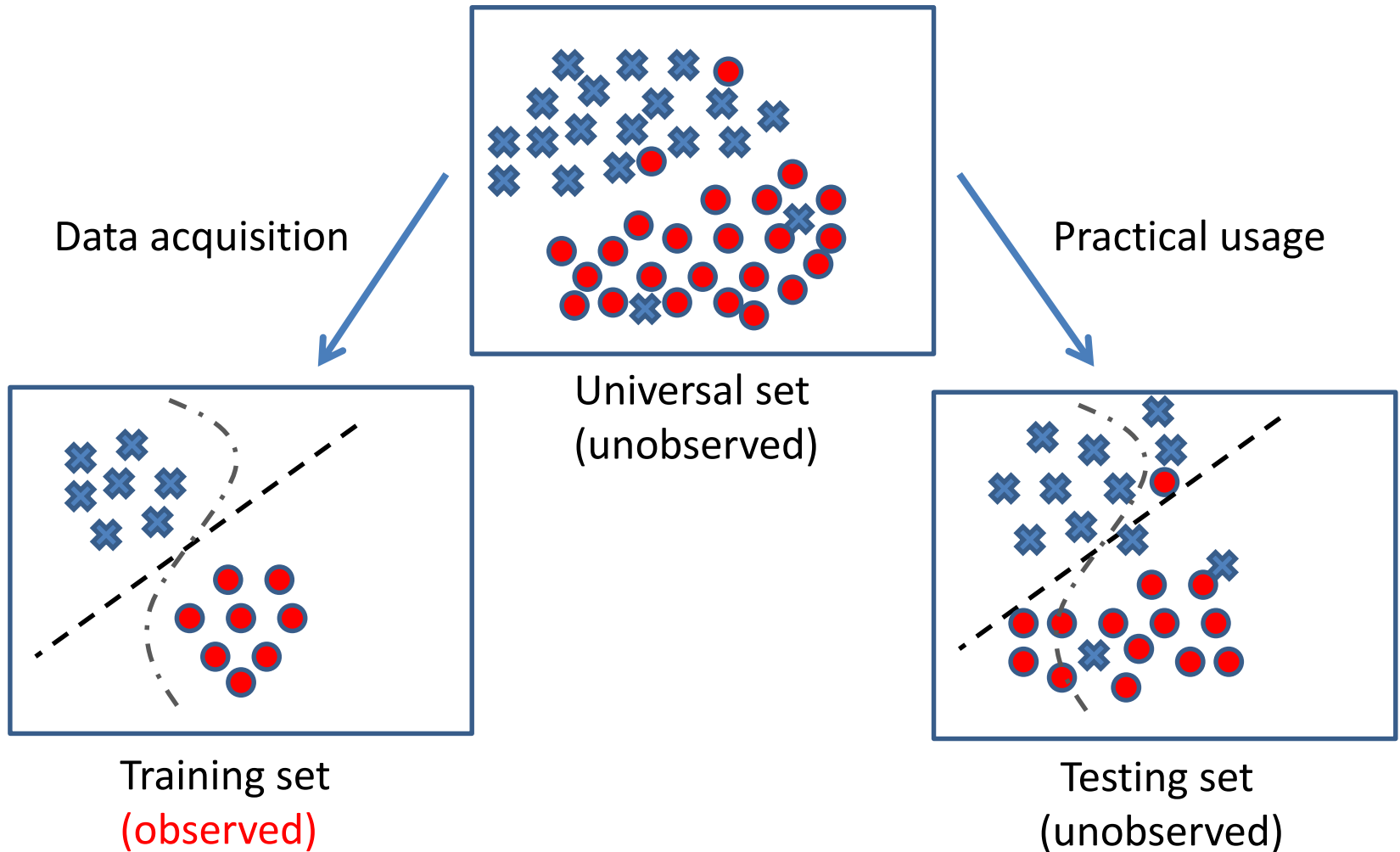
1. A funny thing happened on the way to AI.
  2. Statistics + Algorithms
  3. Fancy Function Fitting
- 
- A branch of **artificial intelligence**, concerned with the **design and development of algorithms** that allow computers to **evolve behaviors** based on empirical data.
  - As intelligence requires knowledge, it is necessary for the computers to **acquire** knowledge.



# Learning system model

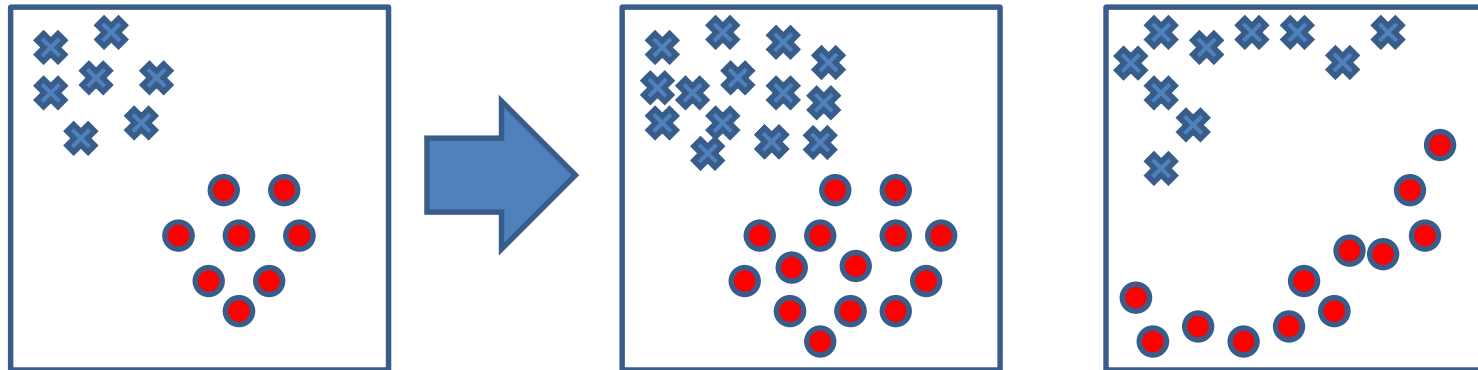


# Training and testing



# Training and testing

- Training is the process of making the system able to learn.
- **No free lunch rule:**
  - Training set and testing set come from the same distribution
  - Need to make some **assumptions** or **bias**
- **Occam's razor:** law of parsimony
  - when presented with competing hypothetical answers to a problem, one should select the one that makes the fewest assumptions.



# Performance

---

- There are several factors affecting the performance:
  - Types of **training** provided
  - The form and extent of initial background **knowledge**
  - The type of **feedback** provided
  - The learning **algorithms** used
- Two important factors:
  - **Modeling**
  - **Optimization**

# Algorithms

---

- The success of machine learning system
  - significantly depends on the algorithms.
- The algorithms control the **search**
  - to find and build the knowledge structures.
- The learning algorithms should
  - extract useful information
  - from training examples.



# Organizing Principle

---

ML algorithm = **representation** +  
**loss function** +  
**optimizer**

$$\min_{r \in R} L(r)$$

- Loss/error function is with respect to data
- Optimizer might not be perfect
  - computationally intractable

# Types of Learning

---

## ◆ supervised $X, y$

- Given an observation  $x$ 
  - what is the best label  $y$ ?
- Given input/output examples, finds mapping
- **Predictive:** What will happen?

# Supervised Learning

## Speech: recognizing dictation



You can dictate your blog or search the web by voice.

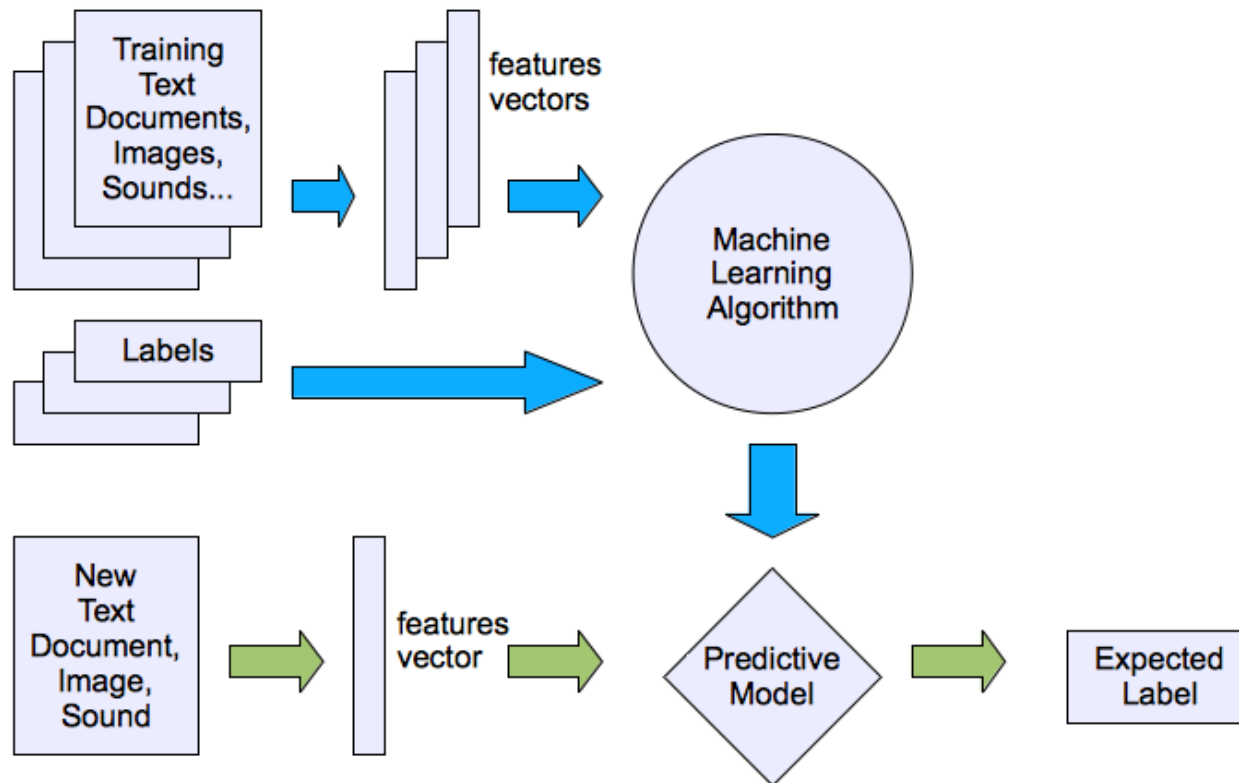
Speech recognition systems are built using (mostly) supervised learning.

# Learning techniques

- Supervised learning categories and techniques
  - **Linear classifier** (numerical functions)
  - **Parametric** (Probabilistic functions)
    - Naïve Bayes, Gaussian discriminant analysis (GDA), Hidden Markov models (HMM), Probabilistic graphical models
  - **Non-parametric** (Instance-based functions)
    - $K$ -nearest neighbors, Kernel regression, Kernel density estimation, Local regression
  - **Non-metric** (Symbolic functions)
    - Classification and regression tree (CART), decision tree
  - **Aggregation**
    - Bagging (bootstrap + aggregation), Adaboost, Random forest

# Machine learning structure

- Supervised learning





# Types of Learning

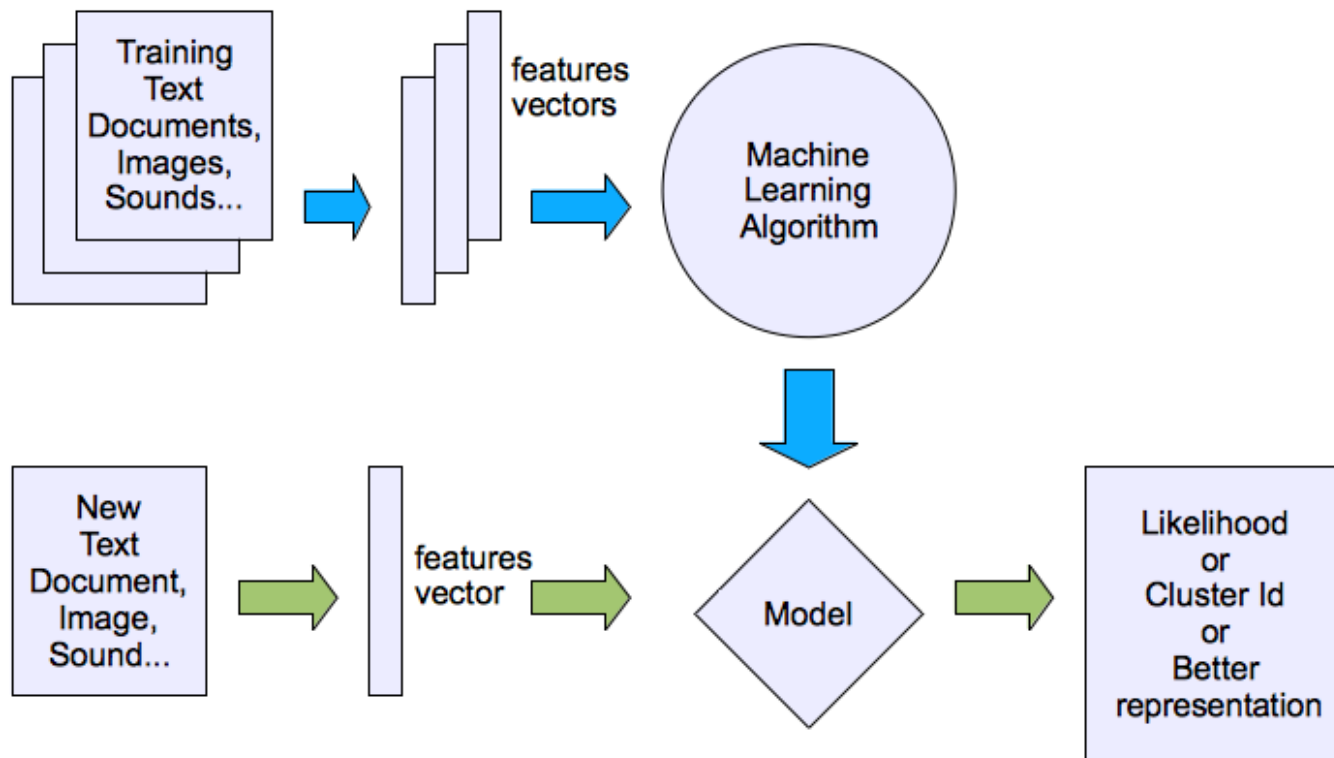
---

## ◆ unsupervised $X$

- Given a set of  $x$ 's
  - cluster or summarize them
- Given data, finds a representation
- **Descriptive:** What happened?

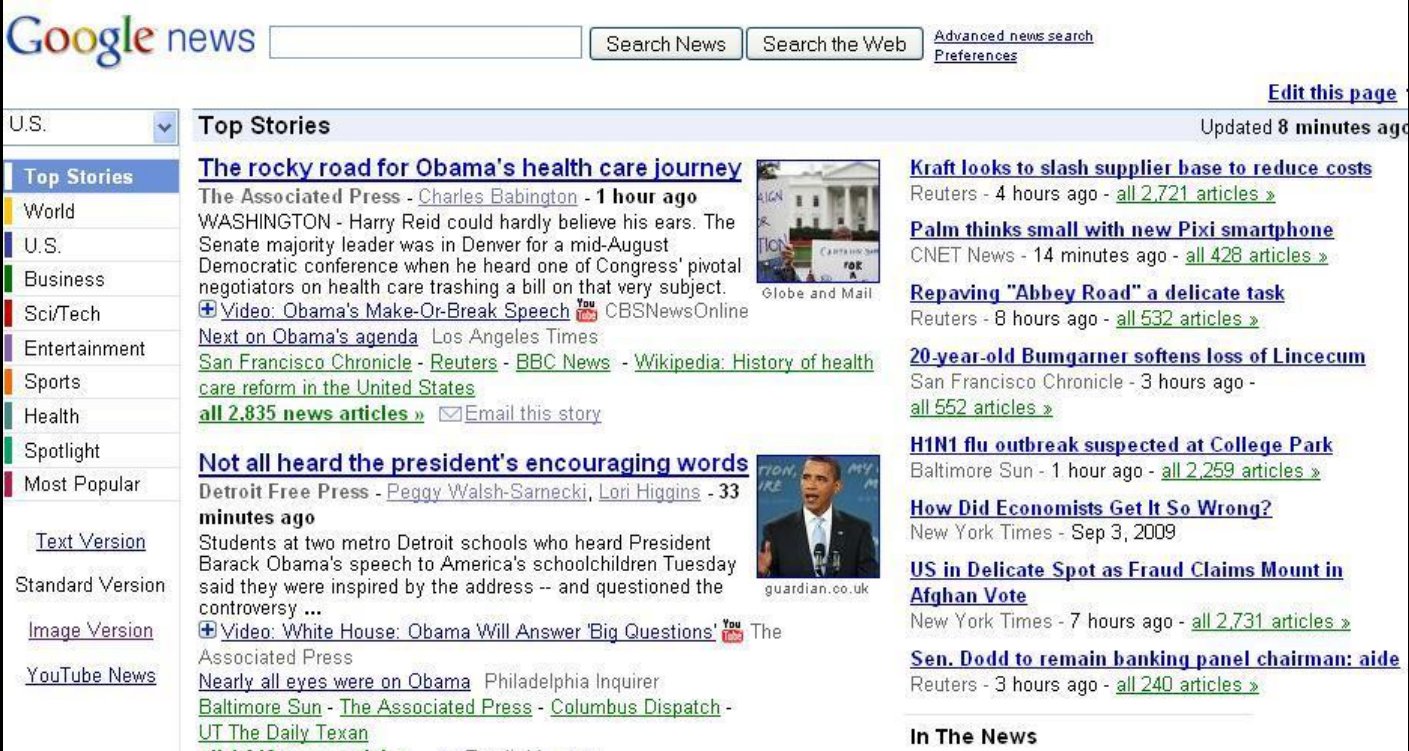
# Machine learning structure

- Unsupervised learning



# Unsupervised Learning

## Text: News digest



The screenshot shows the Google News homepage. At the top, there's a search bar with "Google news" and buttons for "Search News" and "Search the Web". Below the search bar, there's a "U.S." dropdown menu and a "Top Stories" section. The "Top Stories" section is divided into two columns. The left column lists various news categories: Top Stories, World, U.S., Business, Sci/Tech, Entertainment, Sports, Health, Spotlight, and Most Popular. The right column displays several news stories with headlines, source information, and timestamps. The stories include: "The rocky road for Obama's health care journey" (The Associated Press, 1 hour ago), "Kraft looks to slash supplier base to reduce costs" (Reuters, 4 hours ago), "Palm thinks small with new Pixi smartphone" (CNET News, 14 minutes ago), "Repaving 'Abbey Road' a delicate task" (Reuters, 8 hours ago), "20-year-old Bumgarner softens loss of Lincecum" (San Francisco Chronicle, 3 hours ago), "H1N1 flu outbreak suspected at College Park" (Baltimore Sun, 1 hour ago), "How Did Economists Get It So Wrong?" (New York Times, Sep 3, 2009), "US in Delicate Spot as Fraud Claims Mount in Afghan Vote" (New York Times, 7 hours ago), and "Sen. Dodd to remain banking panel chairman: aide" (Reuters, 3 hours ago). Each story includes a small thumbnail image and a link to "all articles" or "Email this story".

Google news [Search News] [Search the Web] [Advanced news search](#) [Preferences](#) [Edit this page](#)

U.S. [v] **Top Stories** Updated 8 minutes ago

**Top Stories**

World

U.S.

Business

Sci/Tech

Entertainment

Sports

Health

Spotlight

Most Popular

[Text Version](#)

[Standard Version](#)

[Image Version](#)

[YouTube News](#)

**The rocky road for Obama's health care journey**  
The Associated Press - Charles Babington - 1 hour ago  
WASHINGTON - Harry Reid could hardly believe his ears. The Senate majority leader was in Denver for a mid-August Democratic conference when he heard one of Congress' pivotal negotiators on health care trashing a bill on that very subject.  
[Video: Obama's Make-Or-Break Speech](#) CBSNewsOnline  
[Next on Obama's agenda](#) Los Angeles Times  
[San Francisco Chronicle](#) - [Reuters](#) - [BBC News](#) - [Wikipedia: History of health care reform in the United States](#)  
[all 2,835 news articles »](#) [Email this story](#)

**Not all heard the president's encouraging words**  
Detroit Free Press - Peggy Walsh-Sarnecki, Lori Higgins - 33 minutes ago  
Students at two metro Detroit schools who heard President Barack Obama's speech to America's schoolchildren Tuesday said they were inspired by the address -- and questioned the controversy ...  
[Video: White House: Obama Will Answer 'Big Questions'](#) The Associated Press  
[Nearly all eyes were on Obama](#) Philadelphia Inquirer  
[Baltimore Sun](#) - [The Associated Press](#) - [Columbus Dispatch](#) - [UT The Daily Texan](#)  
[all 1,012 news articles »](#)

**Kraft looks to slash supplier base to reduce costs**  
Reuters - 4 hours ago - [all 2,721 articles »](#)

**Palm thinks small with new Pixi smartphone**  
CNET News - 14 minutes ago - [all 428 articles »](#)

**Repaving "Abbey Road" a delicate task**  
Reuters - 8 hours ago - [all 532 articles »](#)

**20-year-old Bumgarner softens loss of Lincecum**  
San Francisco Chronicle - 3 hours ago - [all 552 articles »](#)

**H1N1 flu outbreak suspected at College Park**  
Baltimore Sun - 1 hour ago - [all 2,259 articles »](#)

**How Did Economists Get It So Wrong?**  
New York Times - Sep 3, 2009

**US in Delicate Spot as Fraud Claims Mount in Afghan Vote**  
New York Times - 7 hours ago - [all 2,731 articles »](#)

**Sen. Dodd to remain banking panel chairman: aide**  
Reuters - 3 hours ago - [all 240 articles »](#)

**In The News**

Google clusters news stories into groups on the same topic and classifies them into sections like World, Sports, Entertainment, Tech, etc.

This is primarily an example of unsupervised learning.

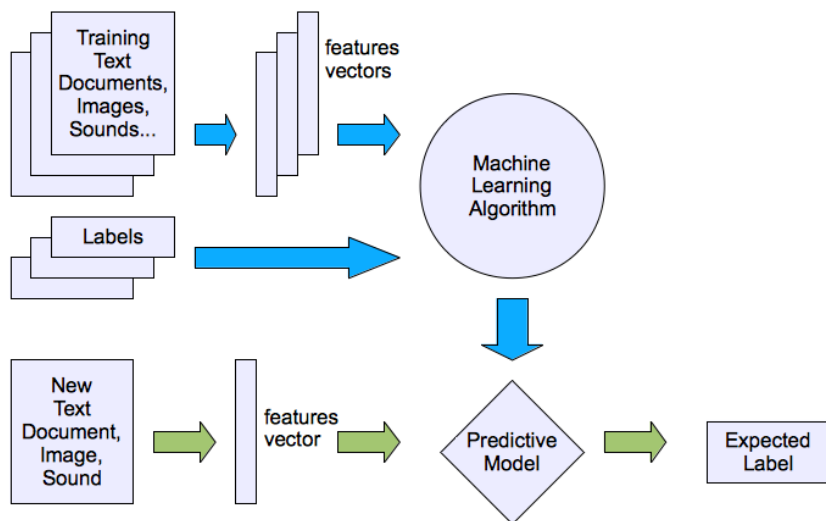
# Learning techniques

---

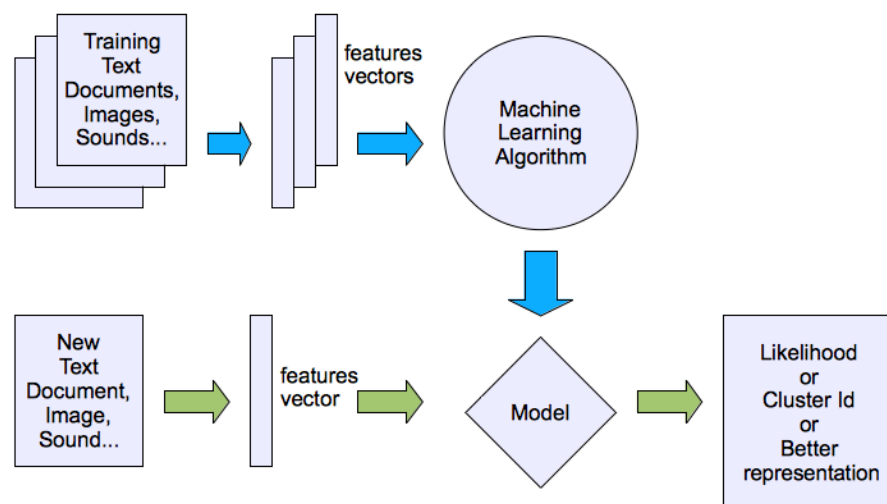
- Unsupervised learning categories and techniques
  - **Clustering**
    - K-means clustering
    - Spectral clustering
  - **Density Estimation**
    - Gaussian mixture model (GMM)
    - Graphical models
  - **Dimensionality reduction**
    - Principal component analysis (PCA)
    - Factor analysis

# Supervised vs. Unsupervised

## Supervised learning



## Unsupervised learning





# Types of Learning

---

## ◆ Semi-Supervised

- can produce considerable improvement in learning accuracy over unsupervised learning

## ◆ Reinforcement

- Translate state to action to maximize reward
- **Prescriptive:** What should we do?

# Types of Learning

## Learning to control robots: Pancake flipping



Complex control policies can be learned using reinforcement learning.

# **Robot Motor Skill Coordination with EM-based Reinforcement Learning**

**Petar Kormushev, Sylvain Calinon,  
and Darwin G. Caldwell**

**Italian Institute of Technology**



00:21 / 03:34



480p



# **Relation to probability**

# Relation to probability

- ◆ ML is (often) modeling a probability distribution
  - Probabilistic reasoning is central to many ML tasks
  - Probability is an extremely useful way of quantifying our beliefs about the state of the world

## ◆ supervised

$X, y$

- $\min |y^{\text{est}}(x) - y|$  - optimization
- $P(y|x)$  - conditional probability estimation

## ◆ unsupervised

$X$

- $P(x)$  - “generative” model

# Generative vs. Discriminative models

---

- ◆ Many of the methods discussed in this course model one of the following probability distributions:
  - Generative :
    - $P(X)$  -- Example: GMM (unsupervised) or
    - $P(X, Y)$  -- Example: Naive Bayes (supervised)
  - Discriminative:
    - $P(Y | X)$  -- Example: Logistic Regression (supervised)

# Generative vs. Discriminative models

---

- ◆ For any generative model : in order to make decisions
  - we can compute a discriminative **posterior** through

$$P(Y \mid X) = P(X, Y) / \sum_{Y'} P(X, Y')$$

- Eg. NB and LR can both lead to the same form of posterior

# Generative vs. discriminative models

- ◆ Any generative model can be used as an unsupervised method.

$$P(X) = \sum_Y P(X, Y)$$

- Since a generative model defines a distribution over  $X$  by **marginalizing** out the class labels.
  - If the class labels are unknown, we can use EM to estimate them!
- ◆ **Graphical models** cover all probability models we discussed.
    - Graphical models are an efficient and intuitive way of encoding a set of independence assumptions about a set of random variables.
    - Once you have learned about graphical models, it's rare to ever talk about probabilities again without them!
    - [Probabilistic Graphical Models - Daphne Koller](#)





# ML is (often) optimization

- ◆ Probability express our beliefs about the state of the universe
- ◆ Optimization is the way to achieve goals
- ◆ **Objective (Loss) functions**
  - captures performance of a model on a given task
  - optimize that objective with respect to some parameters
  - Typical loss functions
    - $f(x)$  is a predictive model that returns a real-valued number

**0–1 (binary)**

$$1(y \neq \text{sign}(f(x)))$$

**Squared**

$$(y - f(x))^2$$

**Exponential**

$$\exp\{-yf(x)\}$$

**log**

$$\log(1 + \exp\{-yf(x)\})$$

**Hinge**

$$\max\{0, 1 - yf(x)\}$$



# ML is (often) optimization

## ◆ Regularization: MLE vs. MAP

- **overfitting** can occur if we optimize our objective on training data too tightly
- To explicitly control the extent to which we trust the training data
- we introduced the concept of **priors** or **regularization**
- If  $\mathcal{D}_X$ ,  $\mathcal{D}_Y$  are the dataset, and  $\theta$  our parameters
- then probabilistically we have

$$\log P(\mathcal{D}_X, \mathcal{D}_Y, \theta) = \log P(\mathcal{D}_X, \mathcal{D}_Y \mid \theta) + \log P(\theta)$$

$$\log P(\mathcal{D}_X, \mathcal{D}_Y, \theta) = -\text{loss}(\theta) + \text{regularizer}(\theta)$$

# Organizing Principle

---

ML algorithm =

{

**representation** +

**loss function** +

**optimizer**

}

# What are we seeking?

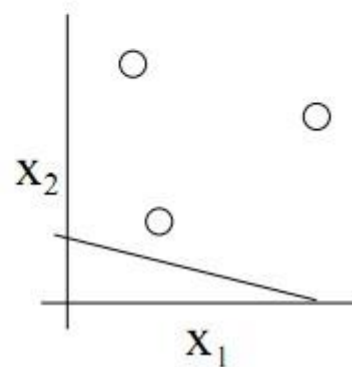
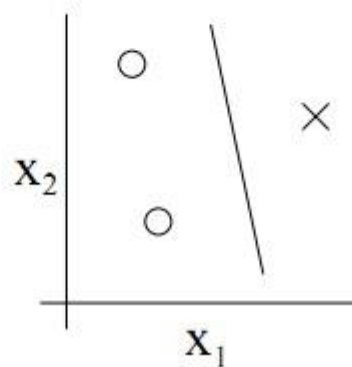
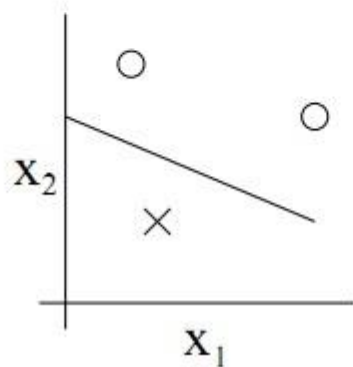
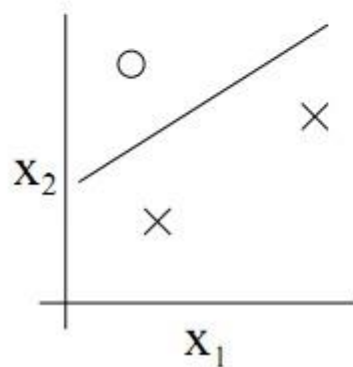
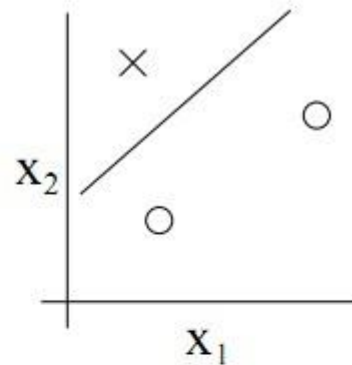
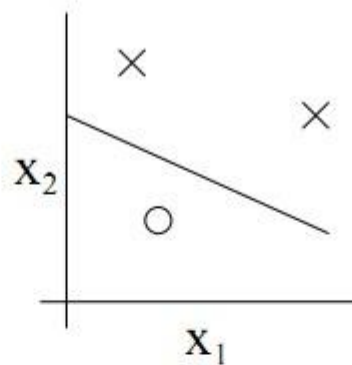
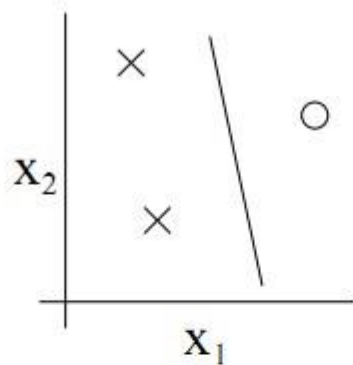
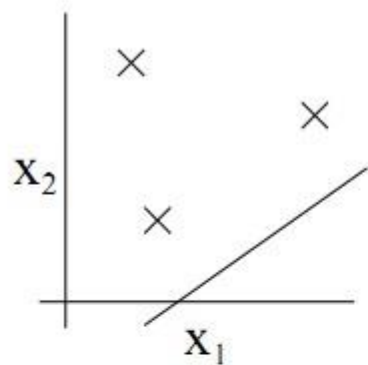
- Supervised: Low E-out or maximize probabilistic terms

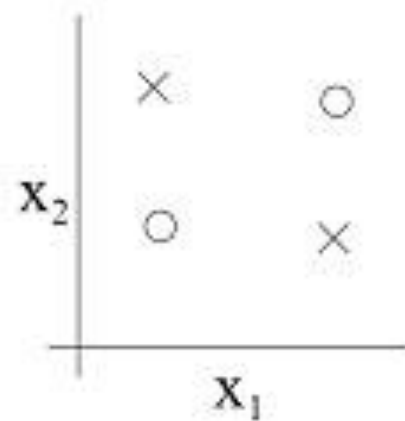
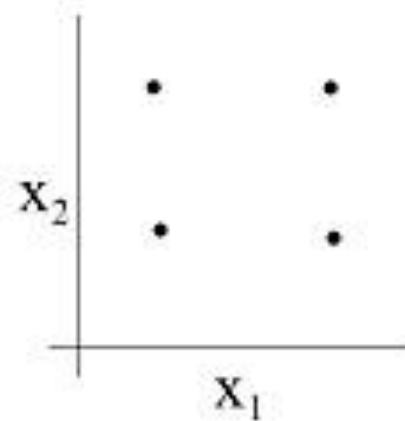
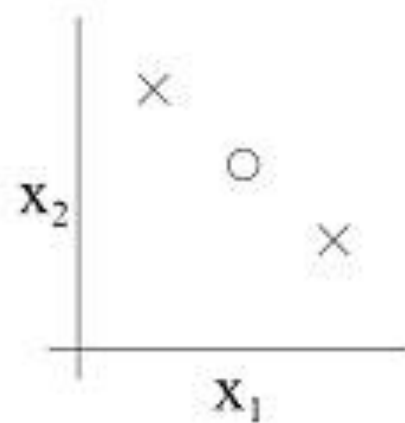
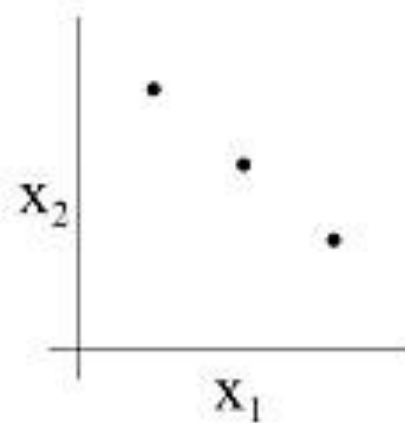
$$error = \frac{1}{N} \sum_{n=1}^N [y_n \neq g(x_n)]$$

E-in: for training set  
E-out: for testing set

$$E_{out}(g) \leq E_{in}(g) \pm O\left(\sqrt{\frac{d_{VC}}{N} \ln N}\right)$$

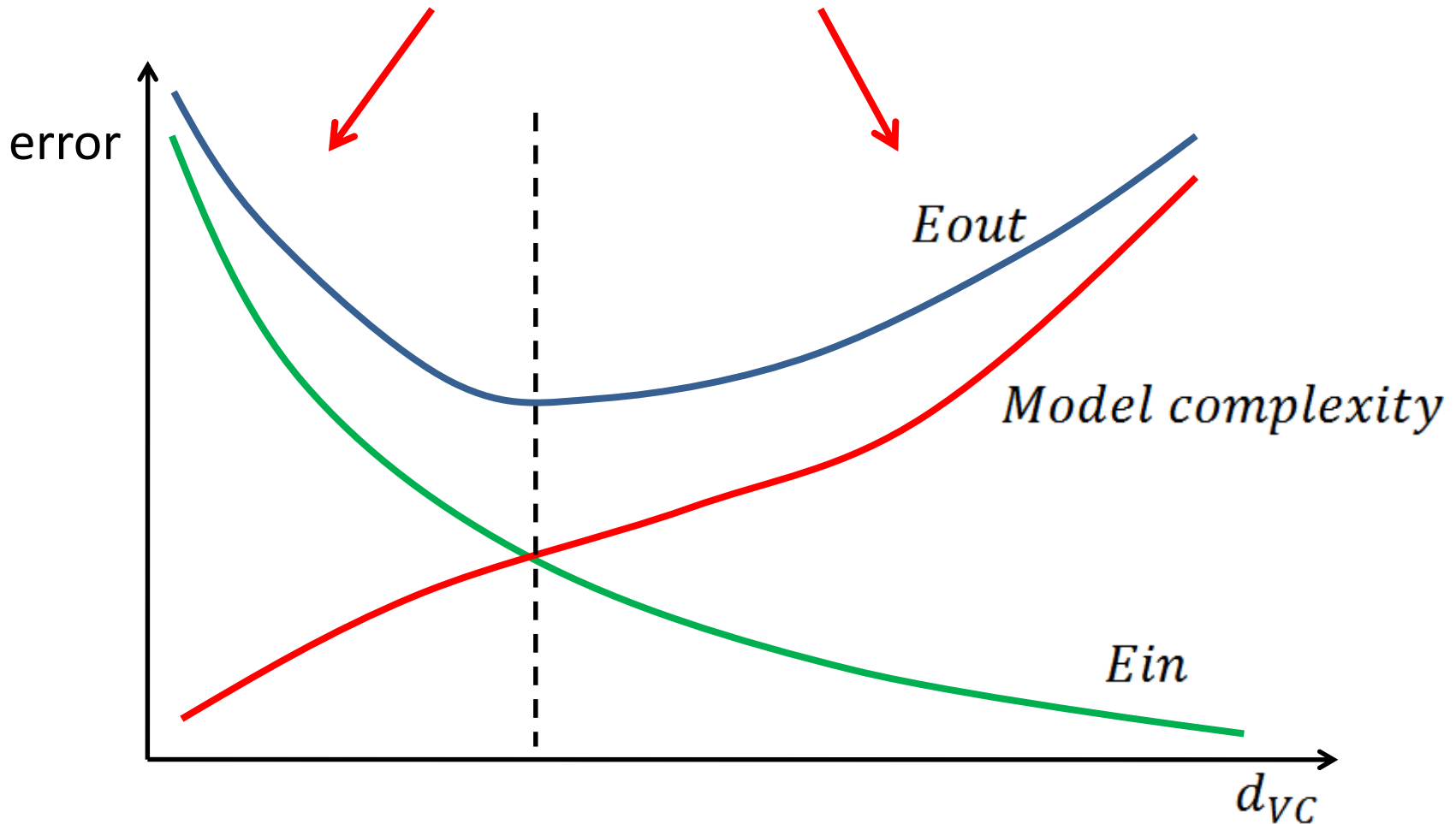
- Unsupervised: Minimum quantization error, Minimum distance, MAP, MLE(maximum likelihood estimation)





# What are we seeking?

Under-fitting VS. Over-fitting (fixed  $N$ )



**Cross validation is of incredibly important!**



# MODEL SELECTION WITH CROSS VALIDATION



00:21 / 03:34



480p



# Cross-validation

---

## ◆ Cross validation is of incredibly important

- It's critical for both training, evaluating, and selecting between different models.
- At this point, you should know how to approach a problem, divide the data into training and test, and compare different algorithms on that problem.

# Conclusion

---

- We have a simple overview of some techniques and algorithms in machine learning.
- Furthermore, there are more and more techniques apply machine learning as a solution.
- In the future, machine learning will play an important role in our daily life.

# What you should know

---

## ◆ Turning a real-world problem into a well-posed ML

problem is often hard

- E.g. generate features/predictors, pick  $X$  and  $y$

## ◆ Unsupervised vs. supervised

- Generative  $P(x)$  vs. conditional  $P(y|x)$  models

