

# 机器学习的力与美

---

陈健生

信息认知与智能系统研究所

清华大学电子工程系

二零一四年五月

- 机器学习
- 三只麻雀
- 学以致用

# 机器学习



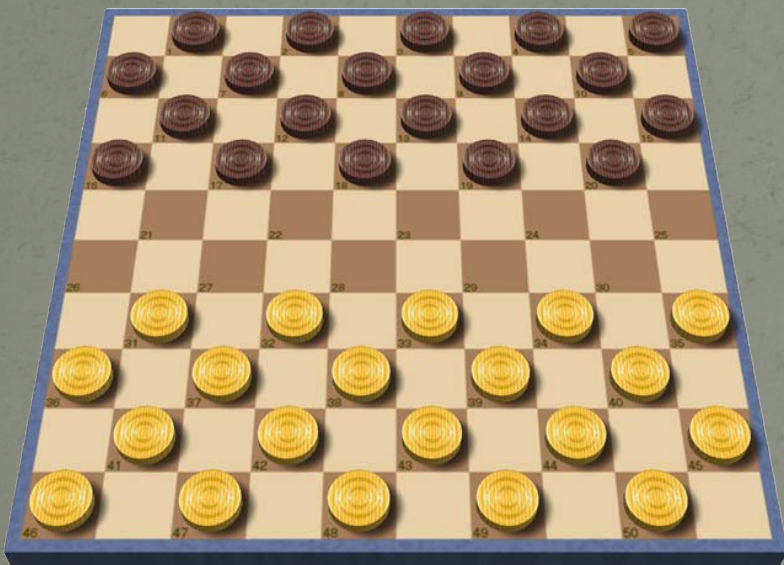
理无专在，学无止境

Can  
Machine  
Learn?

Machine  
Learning!

# 机器学习

“Field of study that gives computers the ability to learn without being explicitly programmed” – A. Samuel, 1959



“Checkers program” (IBM701)

世界上第一个自学习的程序

# 机器学习

"A computer program is said to **learn** from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E" – T.M. Mitchell, 1997

程序从某类经验中**学习**出规律，运用到相关的数据上，在特定的度量下取得更优的预测结果。

# 机器学习

“A breakthrough in machine learning would be worth **ten Microsoft**” - Bill Gates

# 机器学习

- 应用广泛(力)

- 自动控制、信息认知、生物计算、金融分析、数据挖掘 ...

- 研究丰富(美)

- 有监督学习、无监督学习、增强学习、学习理论 ...



LOW UNIVERSITY

YOU = LOVE

AGAINST

FREE  
VARIABLES

FREE  
VARIABLES

SAFER  
DATA  
MINING

PEOPLE  
OVER  
PROFIT

END  
QUALITY  
GAP

BAN  
GENETIC  
ALGORITHMS

LIFT UP  
VECTOR  
MACHINES

Free Variables

FOR WORKERS



# 机器学习

## ➤ 机器学习方法

- 深度学习 (Deep Learning)
- 支持向量机 (Support Vector Machine)
- 流形学习 (Manifold Learning)
- 决策树 (Decision Tree)
- 遗传算法 (Genetic Algorithm)
- 隐马尔科夫模型 (Hidden Markov Model)
- 贝叶斯网络 (Bayesian Networks)
- ... ..

# 机器学习

## ➤ 机器学习方法

➤ 深度学习

➤ 支持向量机

➤ 流形学习

➤ 决策树

➤ 遗传算法

➤ 隐马尔科夫模型 (Hidden Markov Model)

➤ 贝叶斯网络 (Bayesian Networks)

➤ ... ..



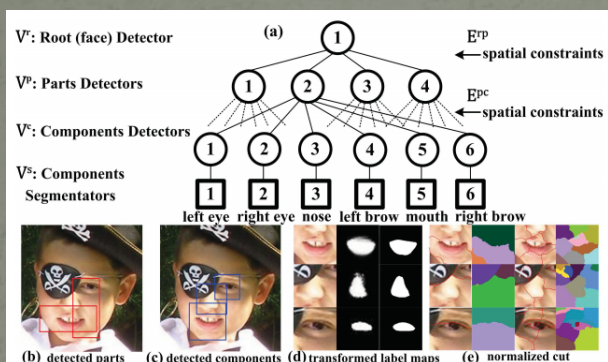
(Decision Tree)

(Genetic Algorithm)

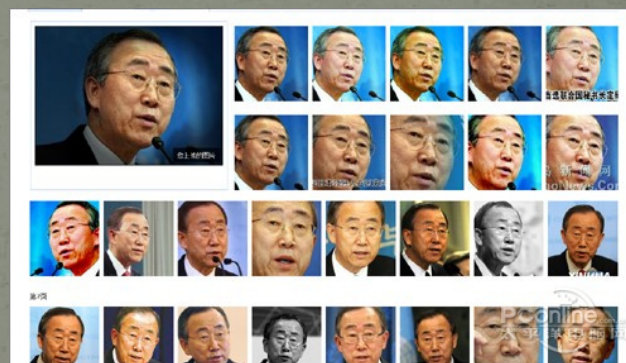
- 机器学习
- 三只麻雀
- 学以致用

# 深度学习

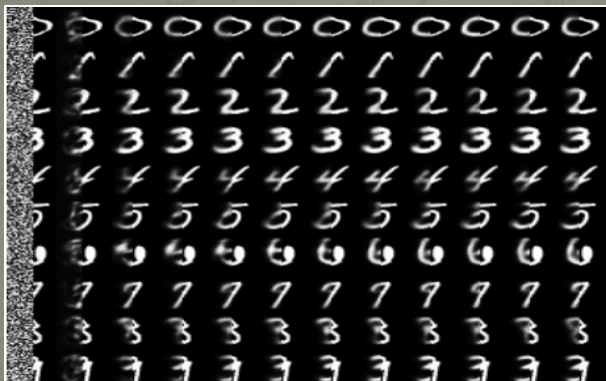
➤ 近年来吸引大量关注 - “Best shot towards AI”



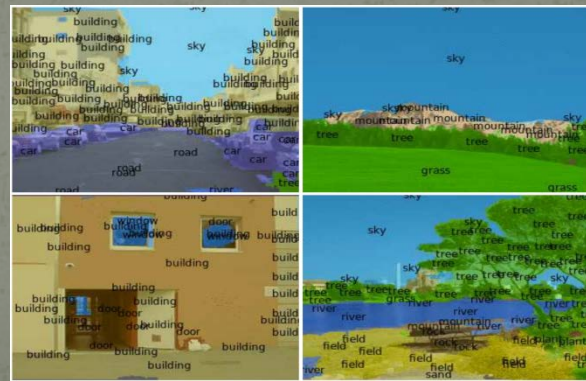
Detection



Retrieval



Recognition

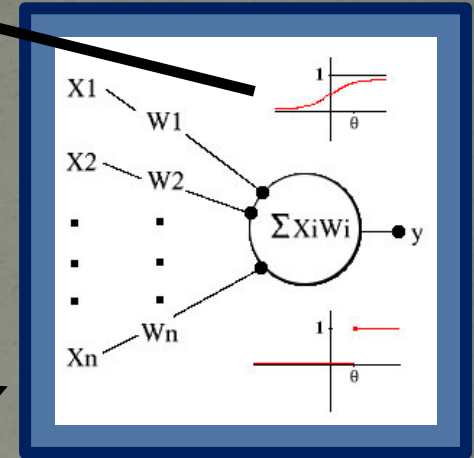
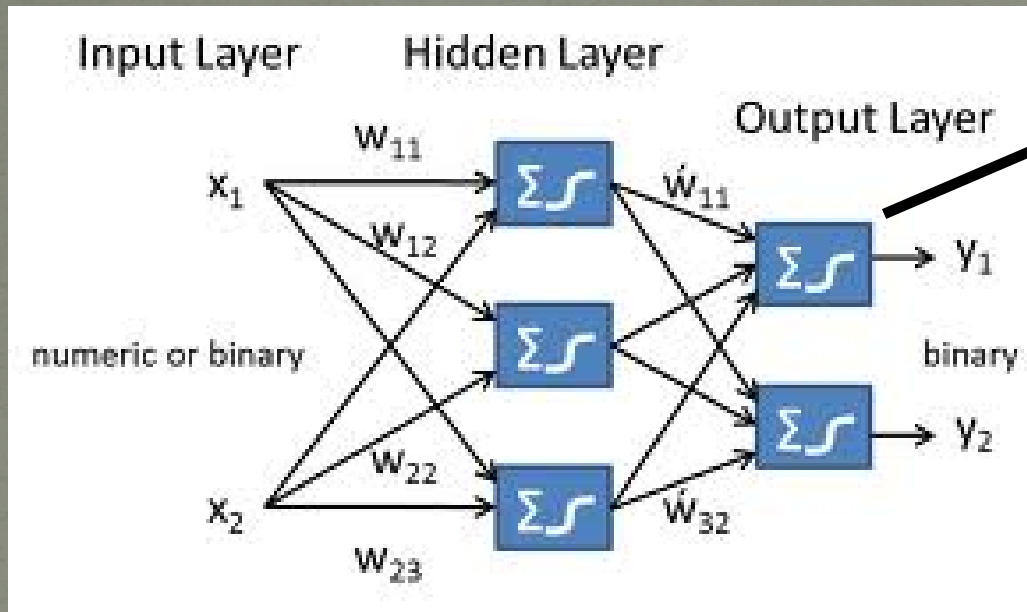


Understanding

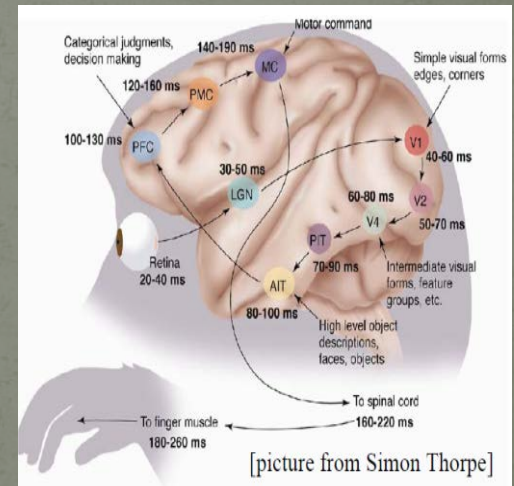
# 深度学习

$$\sigma(a) = \frac{1}{1 + e^{-a}}$$

“浅层”神经网络



$$\vec{Y} = \mathcal{F}(\vec{X} \mid w_{11}, w_{12}, w_{13}, w_{21}, w_{22}, w_{23}, w'_{11}, w'_{12}, w'_{21}, w'_{22}, w'_{31}, w'_{32})$$



# 深度学习

- “网络结构 + 非线性映射” 拟合复杂函数形式
- 采用有监督学习方式训练参数  $W_{11}, W_{12}, W_{13} \dots$

给定训练数据集

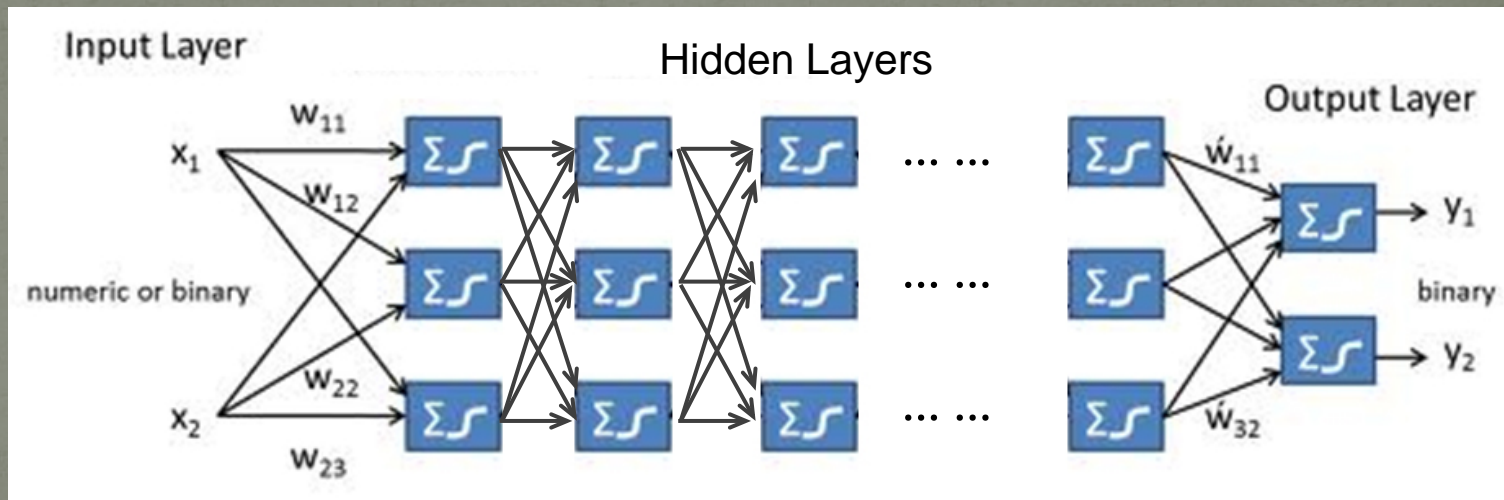
$\vec{Y}_i, \vec{X}_i, i = 1, 2, \dots$

$$\text{minimize } \sum_{i=1,2,\dots} \|\vec{Y}_i - \mathcal{F}(\vec{X}_i)\|$$

- 非凸优化问题，采用逐层梯度下降训练
- “前馈+反向传播”迭代优化
- Very well known by 1990s!

# 深度学习

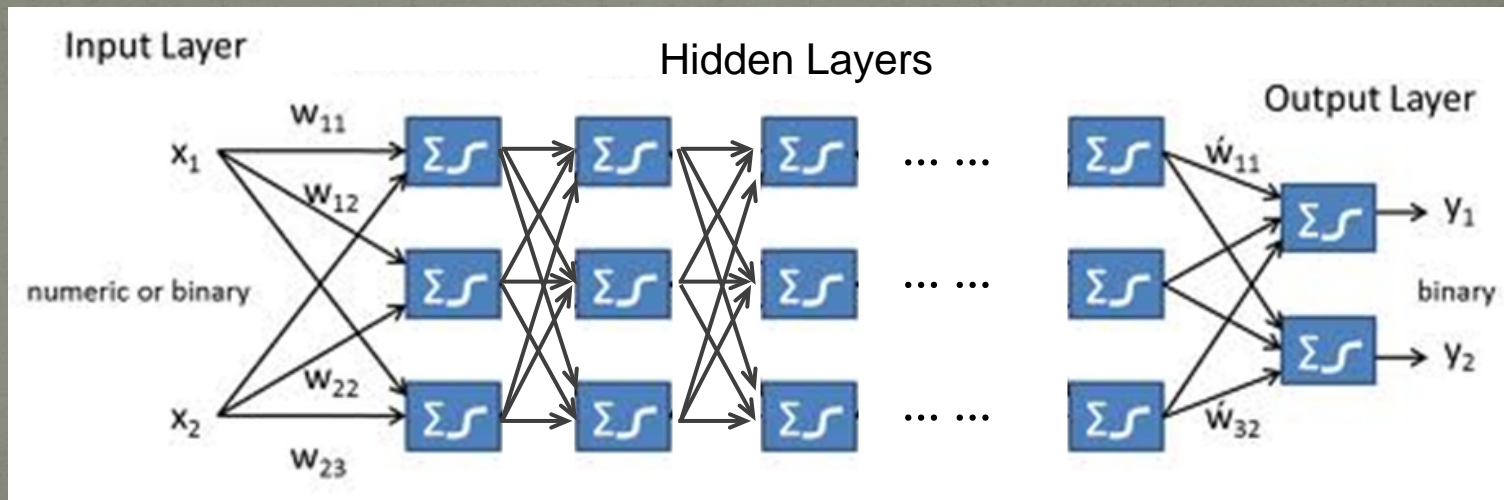
- What's new?
  - pretraining, dropout, ReLU, maxout



“深层”神经网络 – Deep Learning

# 深度学习

- What's new?
  - ~~pretraining~~, dropout, ReLU, maxout
- What really matters?
  - Big Data + Fast Computer



“**深层**”神经网络 – Deep Learning



# 深度学习

## ➤ 优点

- Really works! (MINIST, Netflix, ImageNet, LFW, WSJ, TIMIT, ...)
- 实现简单规范，复用性强
- 易于工业化，very close to \$

## ➤ 缺点

- 非凸优化，易陷入局部最优
- 训练靠经验，结果靠运气
- 理论尚未突破，研究价值??

# 深度学习

力



美



# 支持向量机

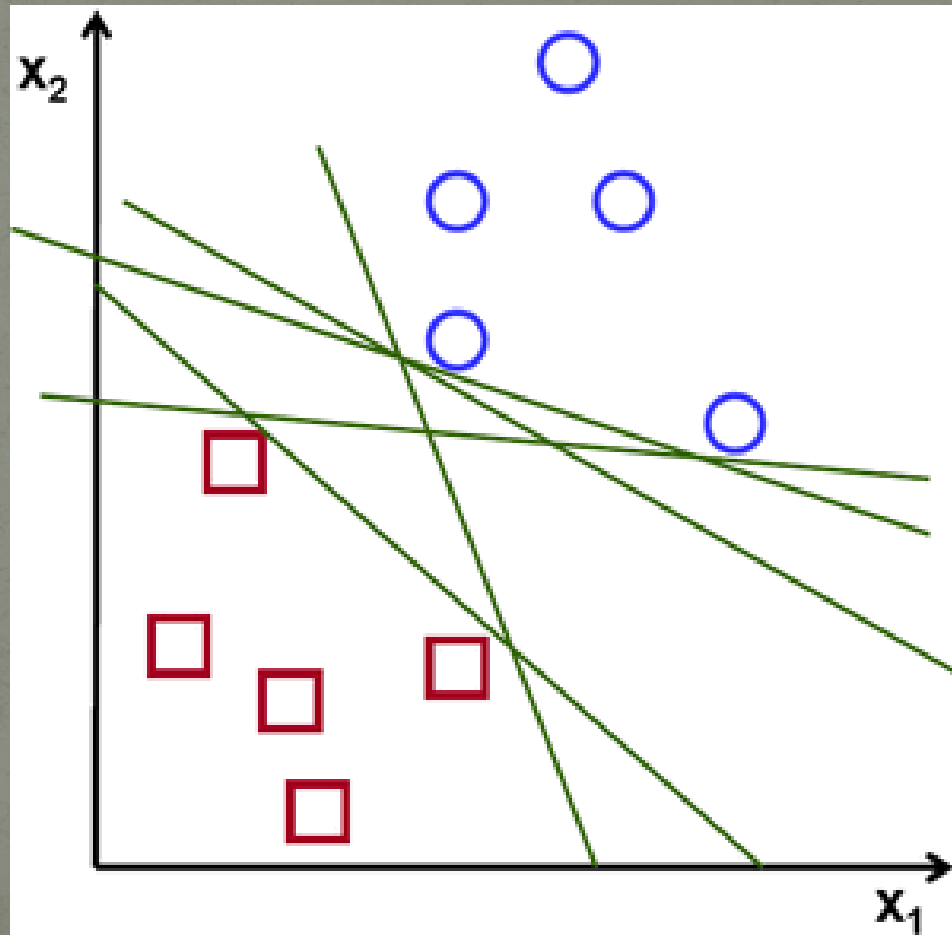


Support Vector Machine

- 提出于90年代早期
- “统治”了机器学习20年
- 很多问题上结果好
- 适合小训练样本情况

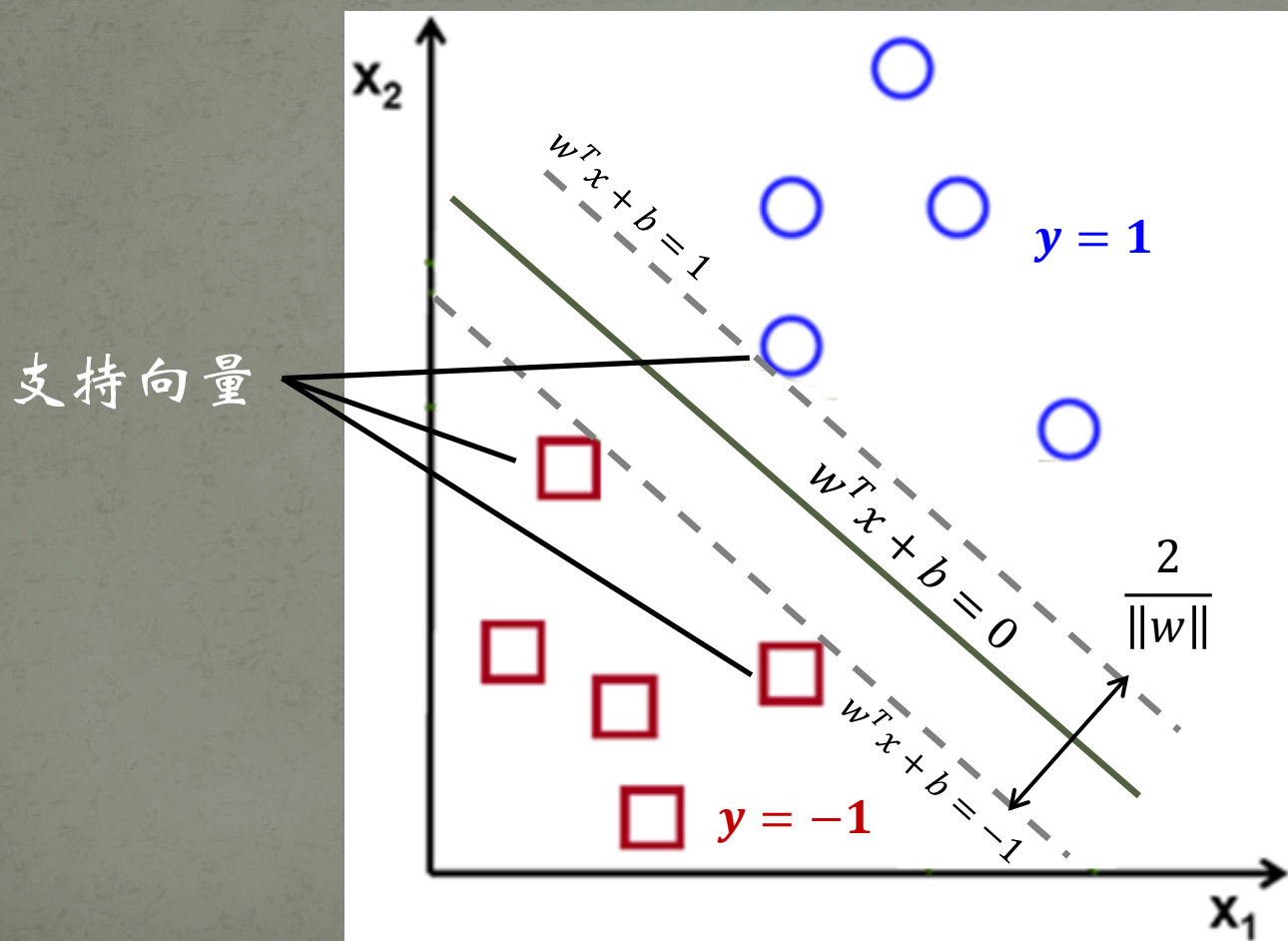
# 支持向量机

## 分类问题



# 支持向量机

## 分类问题



# 支持向量机

- 最大化间隔 or 最小化结构风险

$$\begin{aligned} & \text{minimize } \frac{1}{2} \|w\|^2 \\ & \text{s. t. } y_i (w^T \vec{x}_i + b) \geq 1 \\ & \quad i = 1, 2, 3, \dots \end{aligned}$$

- 凸优化问题，保证全局最优！
- 形式优美，有多种变形
  - Kernel-SVM, Rank-SVM, Struct-SVM, Tran-SVM, ...



# 支持向量机

原问题



对偶问题

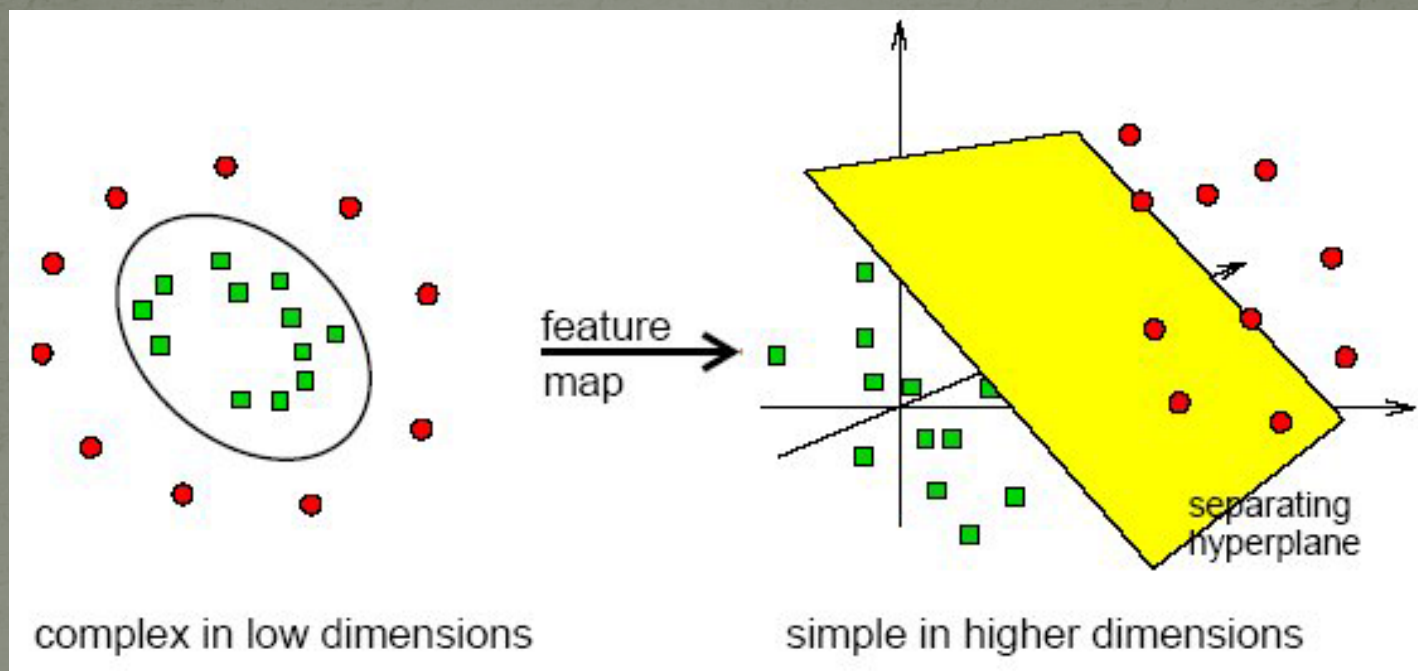
➤ SVM的对偶问题:

$$\text{minimize } \sum_i \alpha_i + \sum_i \sum_j \alpha_i \alpha_j y_i y_j \overrightarrow{x_i x_j}$$

➤ 仅与训练集样本相互间的内积有关

# 支持向量机

## Kernel SVM



Kernel trick: 无需构造高维空间  
只需知道训练样本在其中的内积



# 支持向量机

- 只关心样本之间的相对'大小'关系:

样本对  $x_i x_j$  满足  $y_i > y_j$

- 转化为分类问题:

$$y_i - y_j > 1 \text{ vs } y_j - y_i < -1$$

- 排序支持向量机 Rank SVM

$$\text{minimize } \frac{1}{2} \|w\|^2 + \sum_{i,j} \xi_{i,j}$$

$$\text{s. t. } w^T (\vec{x}_i - \vec{x}_j) \geq 1 - \xi_{i,j}$$

$$\xi_{i,j} > 0$$

# 支持向量机

## ➤ 优点

- 凸优化，保证全局最优
- 实现简单，结果一致
- 构造灵活，变化多端，数学形式优美，研究者的福音
- It works ...

## ➤ 缺点

- 拟合能力较深层结构弱
- 近期在很多结果上输给深度学习，唉...

# 支持向量机

力

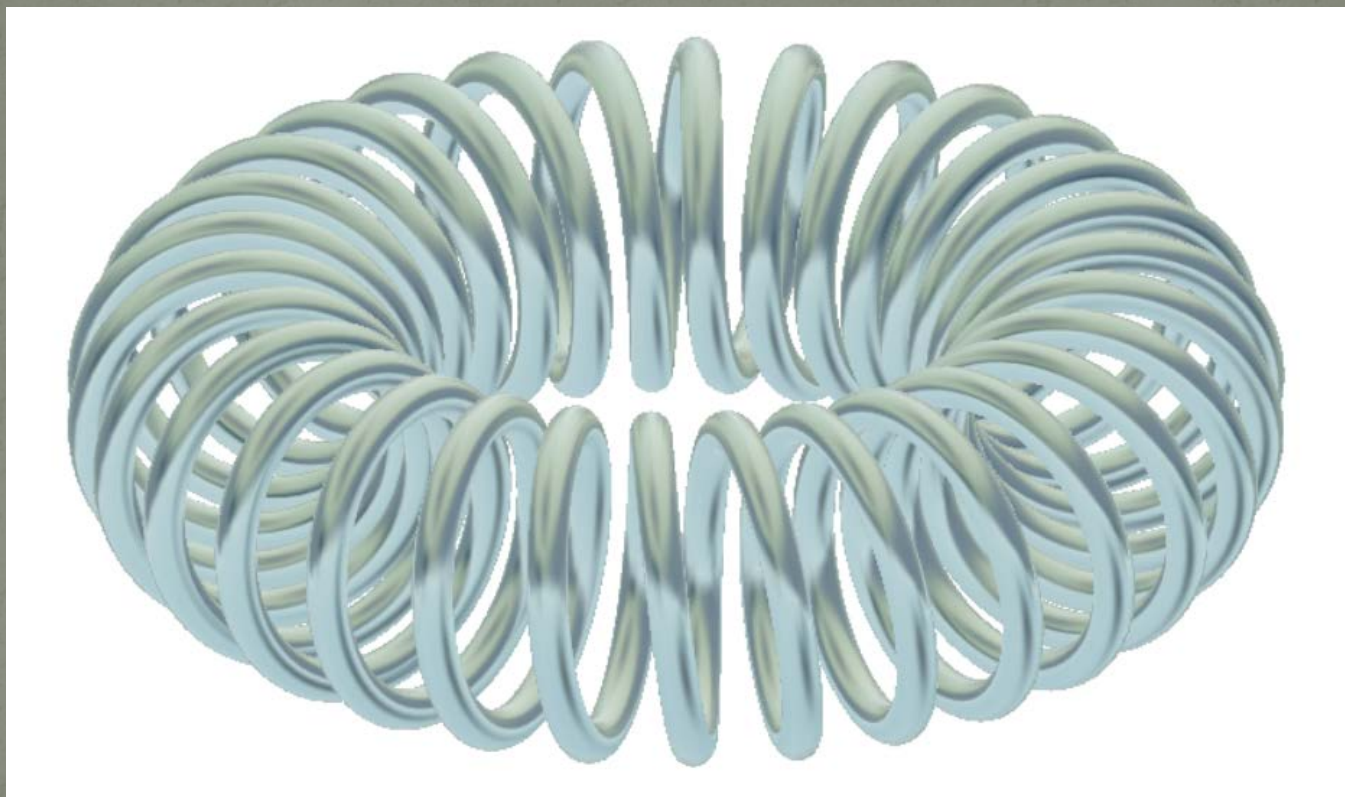


美

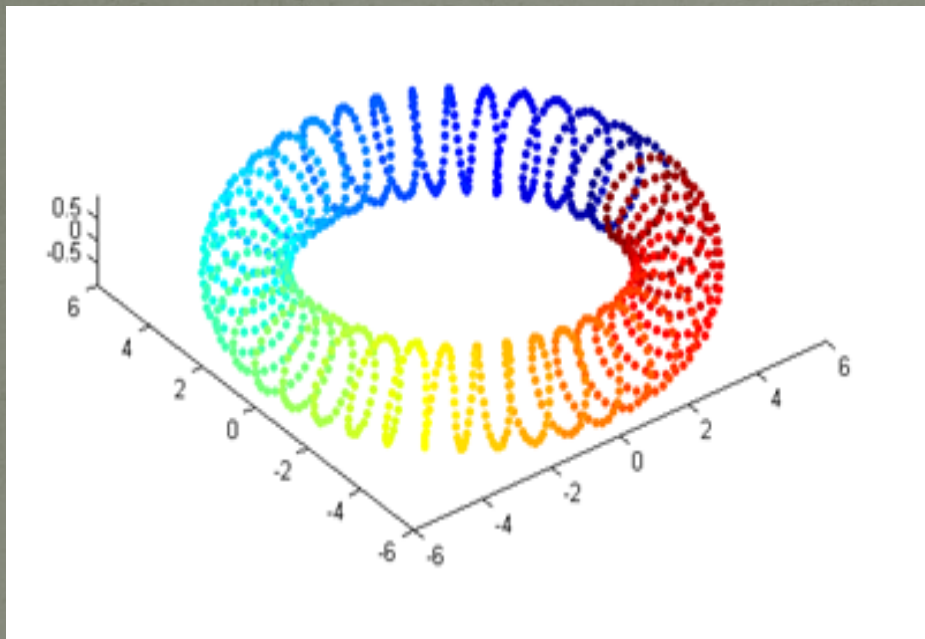


# 流形学习

三维？ 二维？ 一维？



# 流形学习



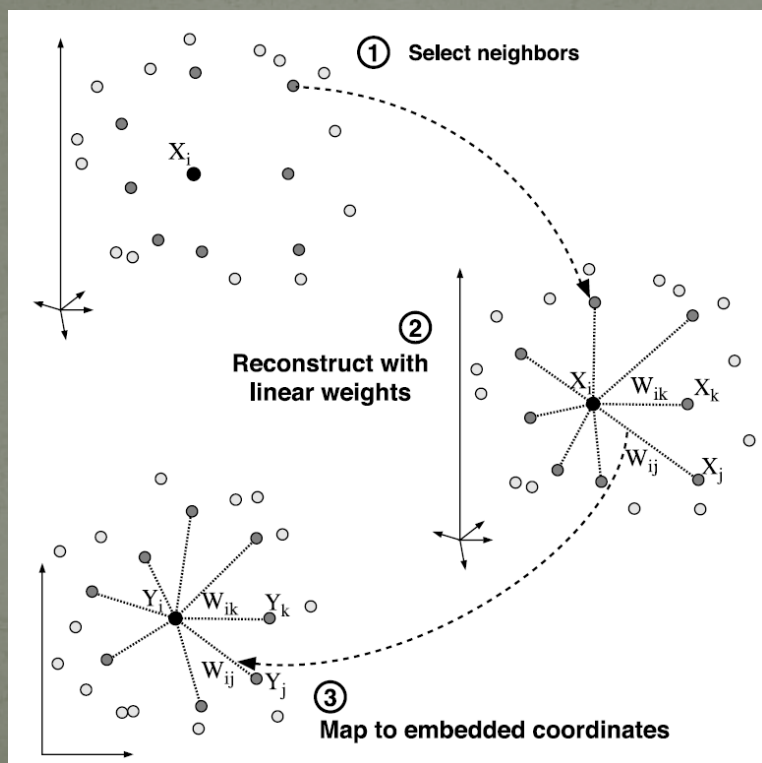
高维空间中的低维形状（流形）

如何在低维空间表示？- 流形学习

# 流形学习

高维空间中的局部结构在低维表示中得以保持

Local Linear Embedding [Science, 2000]



(1) 定位每个数据点  $X_i$  在高维空间中的邻居。

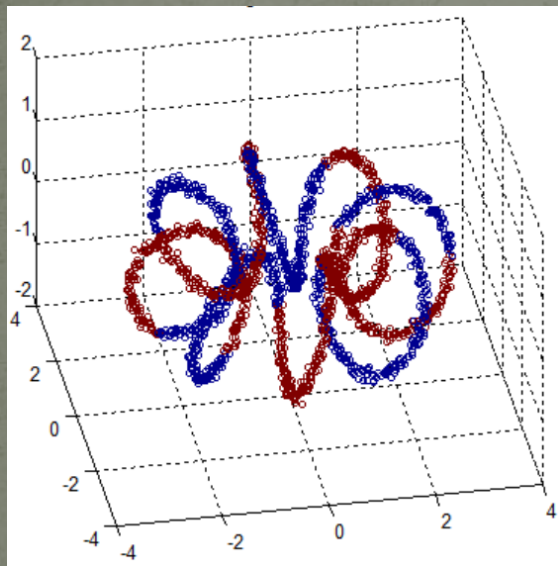
(2) 在高维空间中计算数据点  $X_i$  用其邻居线性组合的最优系数  $W_{ij}$  :

$$\text{minimize } \sum_i \left| \vec{X}_i - \sum_j W_{ij} \vec{X}_j \right|^2$$

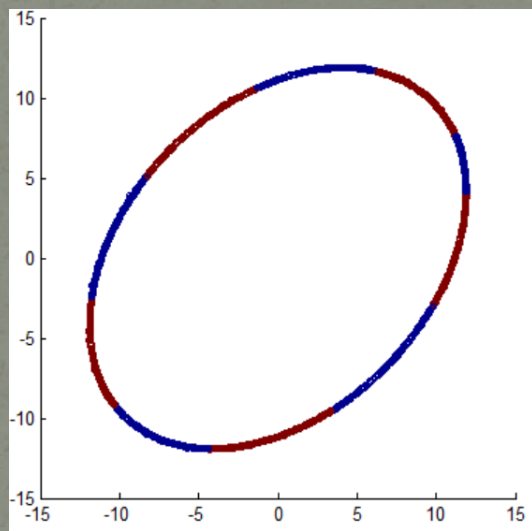
(3) 在低维空间中计算数据点的表示  $Y_i$  使得线性组合系数  $W_{ij}$  得以保持:

$$\text{minimize } \sum_i \left| \vec{Y}_i - \sum_j W_{ij} \vec{Y}_j \right|^2$$

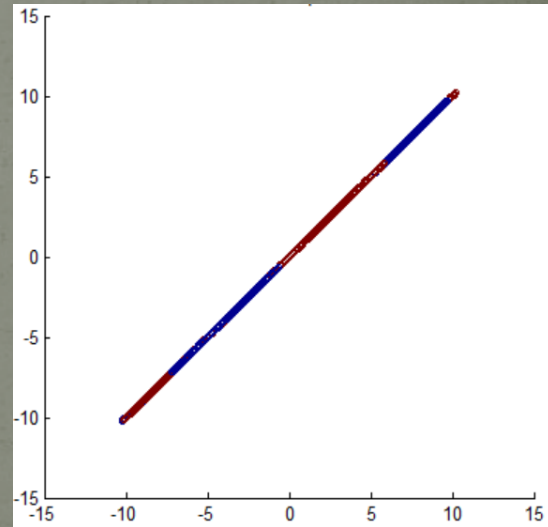
# 流形学习



3维



2维



1维

# 流形学习

## ➤ 优点

- Science!
- 凸优化，保证全局最优
- 可以有效地实现数据降维和结构映射
- It works sometimes ...

## ➤ 缺点

- 要求数据本身真的具备流形性质
- 具有无参性质，使用时效率低
- 结果较难解释，应用性不明确



# 流形学习

力



美



如何使用机器学习工具？

- 机器学习
- 三只麻雀
- 学以致用

# 学以致用

“低分辨率图像的重建与人脸识别”



.....



.....



?

# 学以致用

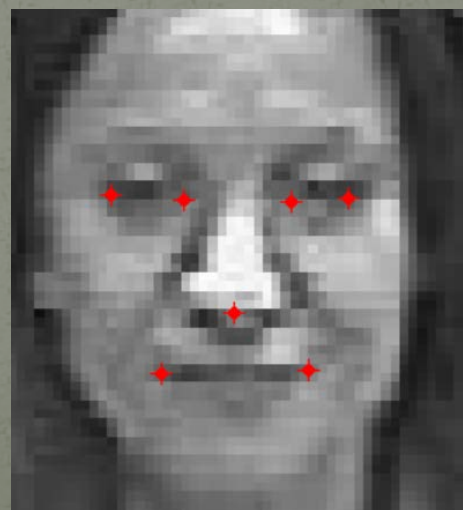
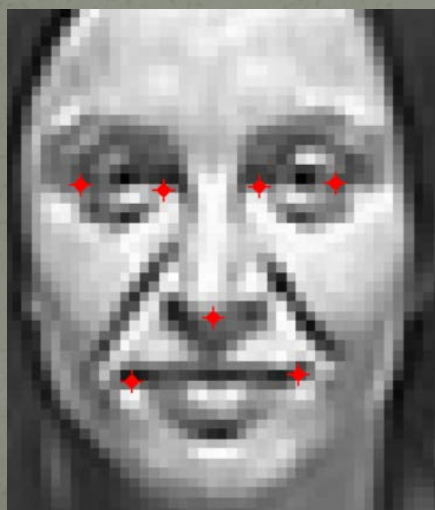
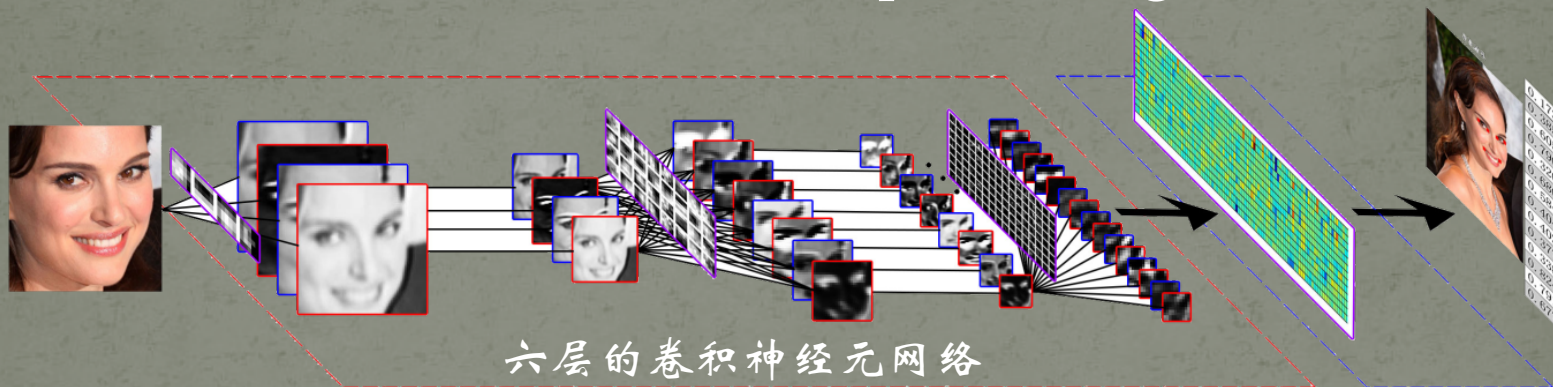


# 学以致用

- 鲁棒的人脸定位与配准 (Deep Learning)
- 自动选择质量较高的人脸 (Rank SVM)
- 提高人脸的图像质量 (Manifold Learning)

# 学以致用

## 鲁棒的人脸定位与配准(Deep Learning)



# 学以致用

- 自动选择质量较高的人脸 (Rank SVM)



低质量



高质量





It is never too late to learn ...



# 参考文献

[Roweis'00] "Nonlinear dimensionality reduction by locally linear embedding"

[Vapnik'95] "Support-vector networks"

[Boser'92] "A training algorithm for optimal margin classifiers"

[Hinton'06] "A fast learning algorithm for deep belief nets"

[Grgic'11] "Scface: surveillance cameras face database"