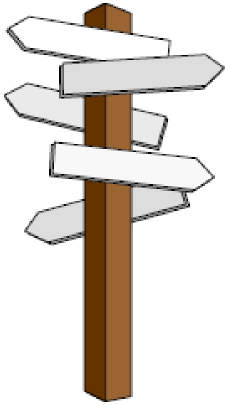
A Brief Introduction to **Machine Learning**
（机器学习简介）

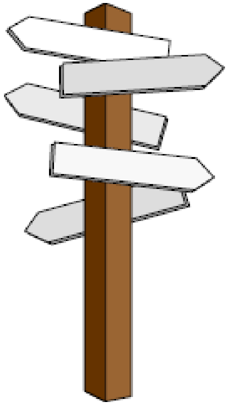
Road Map

- The Concept of Machine Learning
- Learners（学习器）
- Learning Paradigms（学习范式）
- Resources
- Summary



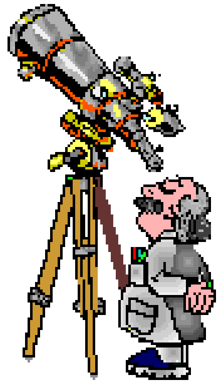
Road Map

- The Concept of Machine Learning
- Learners（学习器）
- Learning Paradigms（学习范式）
- Resources
- Summary



Machine Learning?

- No fixed definition
  - “利用经验改善系统自身的性能” [Mitchell, Book97]
- “Experience” is usually expressed as data in computers
- Main tasks: to learn knowledge from data and make predictions
- DMP:D(Training Data)、M(Model)、P(Prediction)



Play Tennis

Day	Outlook	Temperature	Humidity	Wind	Play Tennis?
1	Sunny	Hot	High	Light	*No*
2	Sunny	Hot	High	Strong	*No*
3	Overcast	Hot	High	Light	*Yes*
4	Rain	Mild	High	Light	*Yes*
5	Rain	Cool	Normal	Light	*Yes*
6	Rain	Cool	Normal	Strong	*No*
7	Overcast	Cool	Normal	Strong	*Yes*
8	Sunny	Mild	High	Light	*No*
9	Sunny	Cool	Normal	Light	*Yes*
10	Rain	Mild	Normal	Light	*Yes*
11	Sunny	Mild	Normal	Strong	*Yes*
12	Overcast	Mild	High	Strong	*Yes*
13	Overcast	Hot	Normal	Light	*Yes*
14	Rain	Mild	High	Strong	*No*
15	Sunny	Hot	Normal	Strong	*Yes*/*No*?

What to Learn?

n*d matrix　　d dimensions　　　　　　label

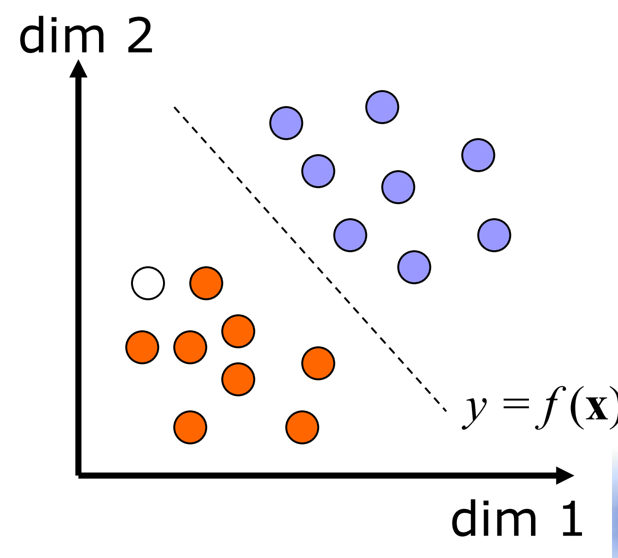
$$\mathbf{X} = \begin{cases} \mathbf{x}_1\text{: instance 1} & \cdots \\ \mathbf{x}_2\text{: instance 2} & \cdots \\ \mathbf{x}_3\text{: instance 3} & \cdots \\ \vdots & \vdots \\ \mathbf{x}_n\text{: instance n} & \cdots \end{cases} \qquad \mathbf{Y} = \begin{cases} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{cases}$$

given a new instance $\mathbf{x}$: ⋯

predict the label y: or ?

Learn a decision function

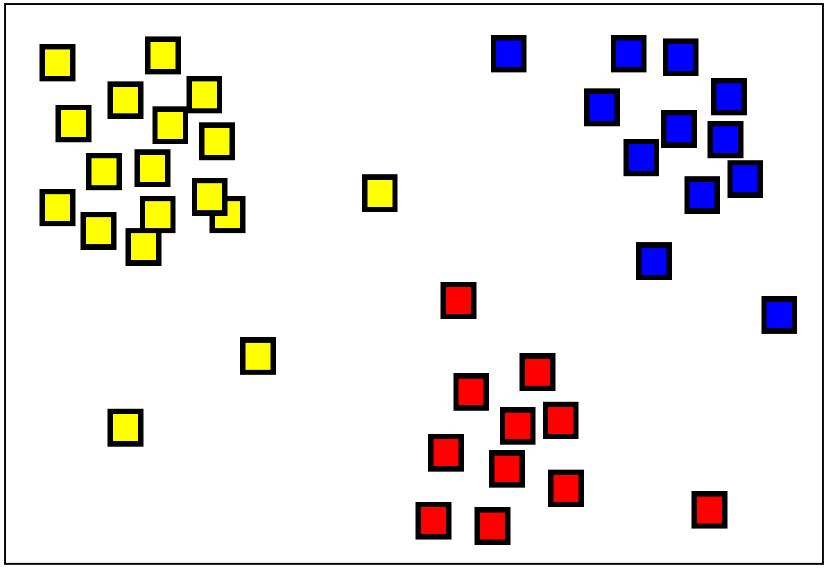
$$y = f(\mathbf{x})$$



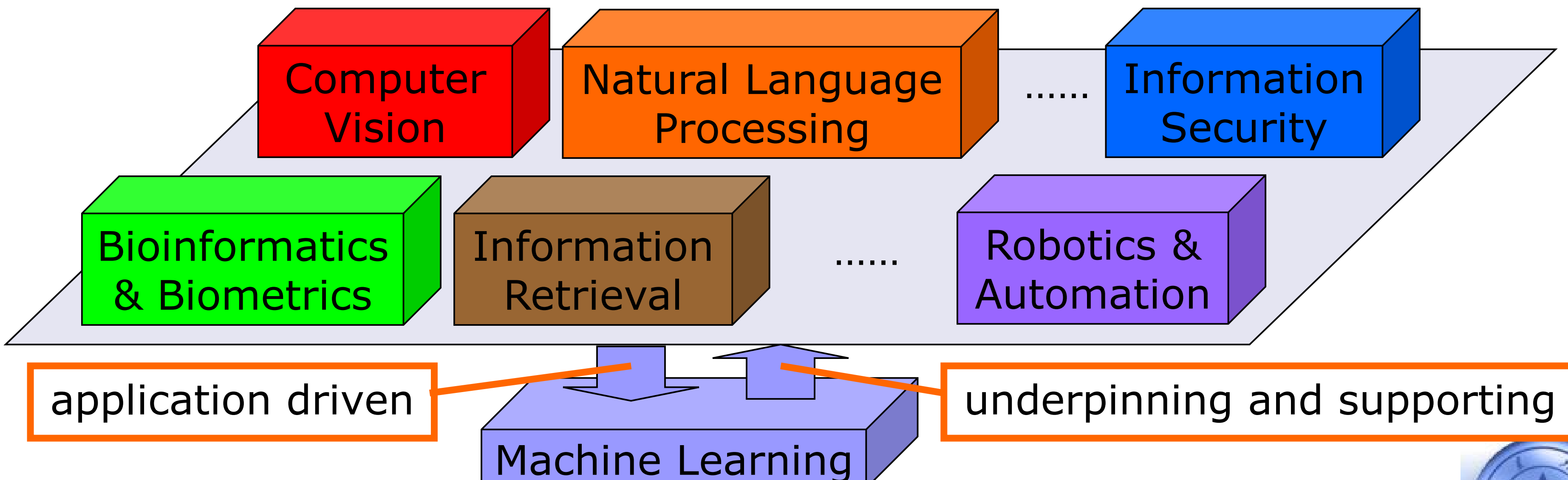
Supervised vs. Unsupervised

- Supervised learning
  - $(\mathbf{X}=\{\mathbf{x}_1,\ldots,\mathbf{x}_n\}, \mathbf{Y}=\{y_1,\ldots,y_n\})$ known, learn $y=f(\mathbf{x})$
  - $\boxed{\text{Classification}}$: y_i is a discrete class label
  - Regression: y_i is a continuous value

- Unsupervised learning
  - $\mathbf{X}$ known, $\mathbf{Y}$ unknown
  - Clustering analysis
  - Anomaly detection

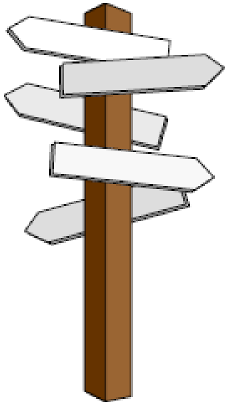


Applications—Machine Learning+



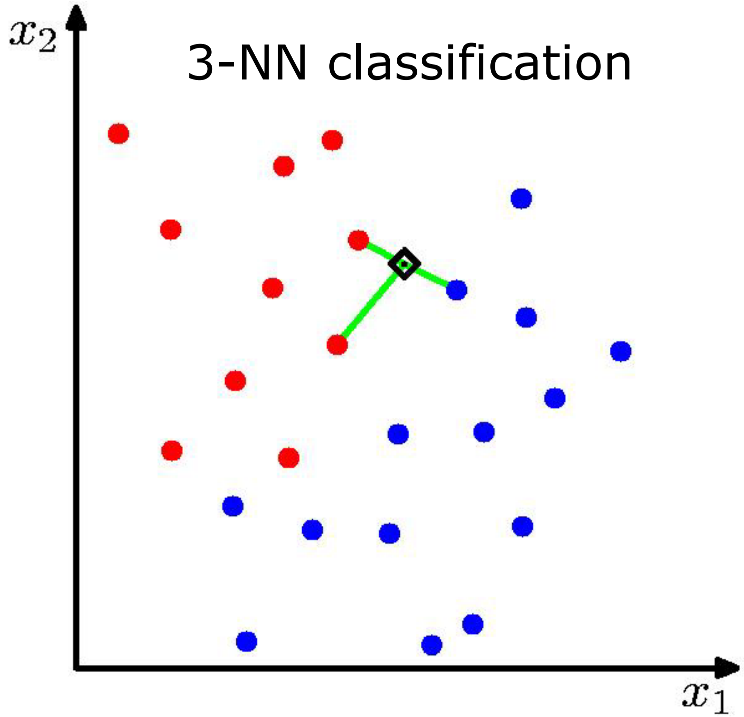
Road Map

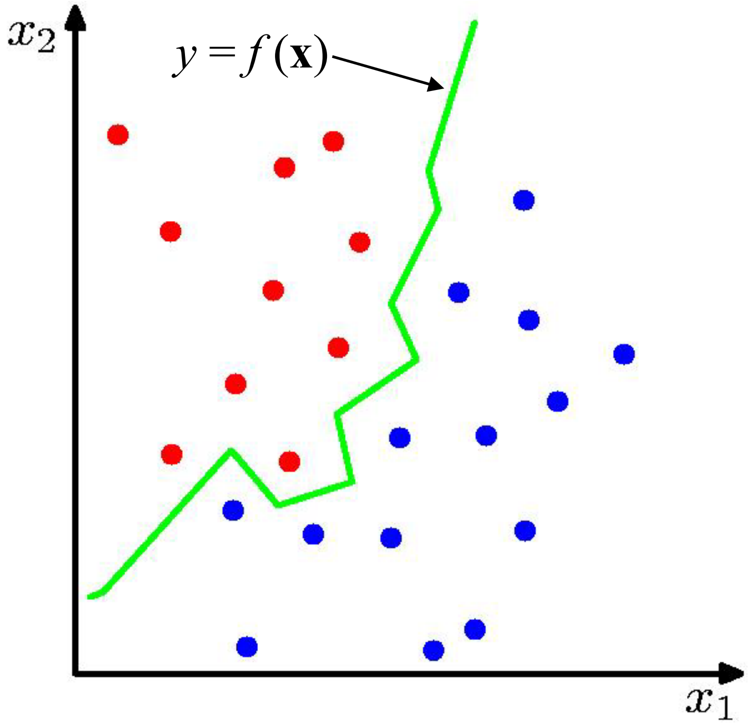
- Machine Learning
- Learners
- Learning Paradigms
- Resources
- Conclusion



k-Nearest Neighbors (k-NN)

- For each unknown instance
  - find its k nearest neighbors: distance metric
  - pick the class label with the most votes





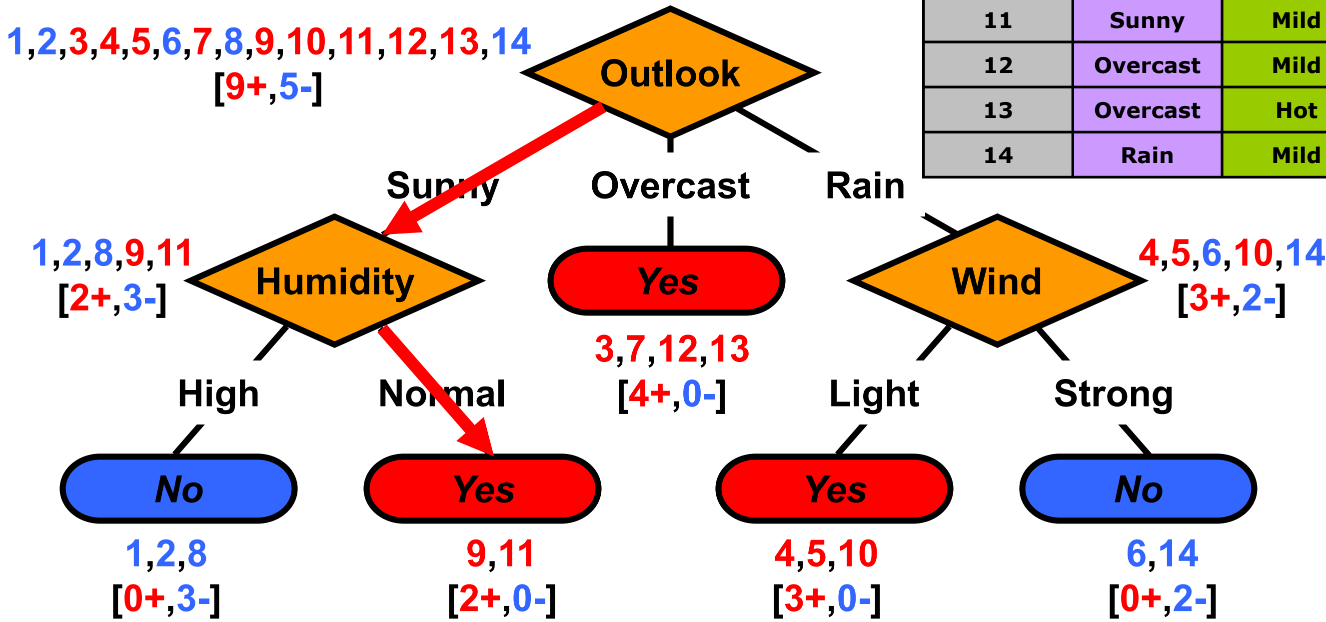
k-NN 特点

- 思想很简单，但经常很有效，至今在很多问题上，仍作为分类器使用
- 没有训练（建模）的过程：without training
- 属于非线性分类器
- 但，当标记样本数量很大、待处理对象维度很高时，计算复杂度很大；
- 也有很多值得研究的空间，如：如何降低计算复杂度、加权近邻法（距离远，权重小）尽管原理简单，

Decision Tree

- Play Tennis

Day	Outlook	Temp.	Humidity	Wind	Play?
15	Sunny	Hot	Normal	Strong	*Yes*/*No?*



Day	Outlook	Temp.	Humidity	Wind	Play?
1	Sunny	Hot	High	Light	*No*
2	Sunny	Hot	High	Strong	*No*
3	Overcast	Hot	High	Light	*Yes*
4	Rain	Mild	High	Light	*Yes*
5	Rain	Cool	Normal	Light	*Yes*
6	Rain	Cool	Normal	Strong	*No*
7	Overcast	Cool	Normal	Strong	*Yes*
8	Sunny	Mild	High	Light	*No*
9	Sunny	Cool	Normal	Light	*Yes*
10	Rain	Mild	Normal	Light	*Yes*
11	Sunny	Mild	Normal	Strong	*Yes*
12	Overcast	Mild	High	Strong	*Yes*
13	Overcast	Hot	Normal	Light	*Yes*
14	Rain	Mild	High	Strong	*No*

Keys for constructing a DT:
(1) splitting dimension
(2) splitting value (for continuous dims)

ID3 **[Quinlan, MLJ86]**
CART **[Brieman et al, Book84]**
C4.5 **[Quinlan, Book93]**

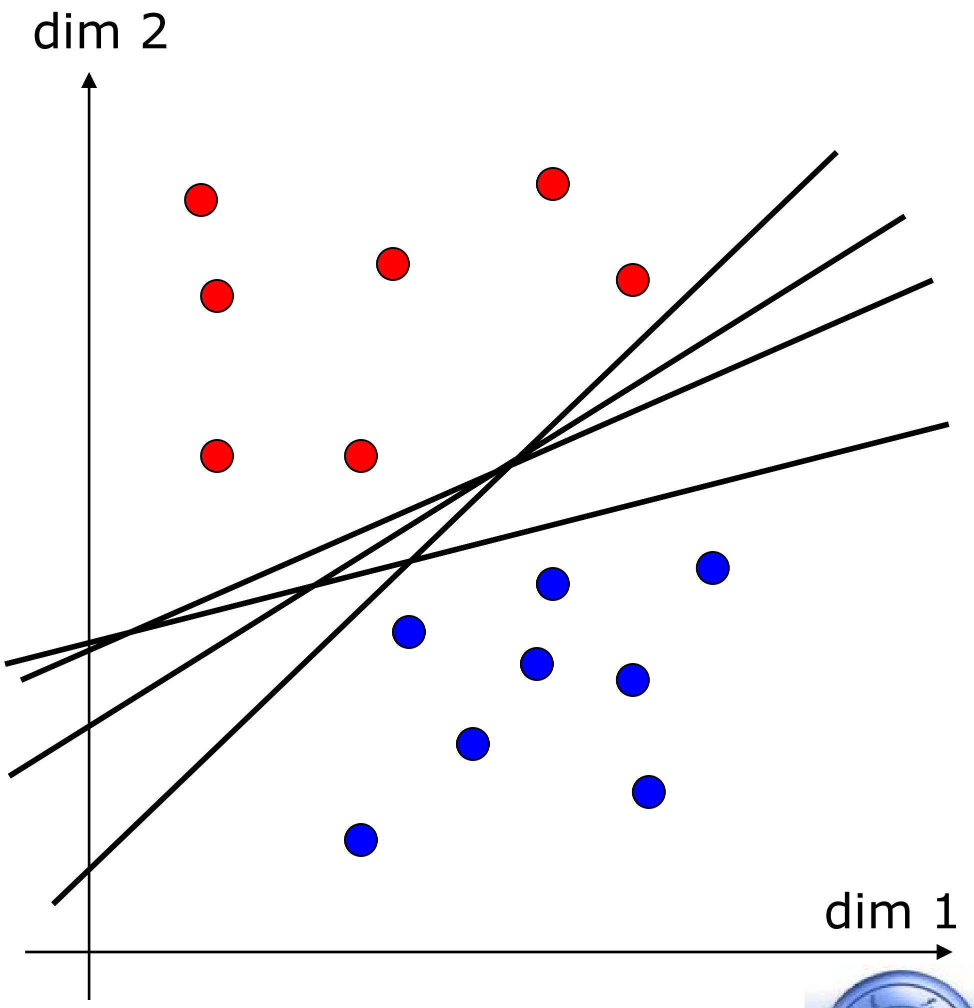
Decision Tree
特点

- 得到的是一组规则集
- 决策过程具有良好的可理解性
- 对分类问题，在解决每类呈现多决策域分布、且交错分布的问题时，具有独特的优势
- 对于单一因素（特征）即可决定预测结果的情况，基于统计机器学习有时会出现的某些问题。如天气预测，如果湿度低到一定的值，就决定了肯定不会下雨；但在统计机器学习的思路下，如果其他多项条件都符合下雨条件，湿度因素就会被平均掉、忽略，从而会得出会下雨的错误预测。用决策树，则可以避免此类问题的出现。

dim 1

Support Vector Machine (SVM)

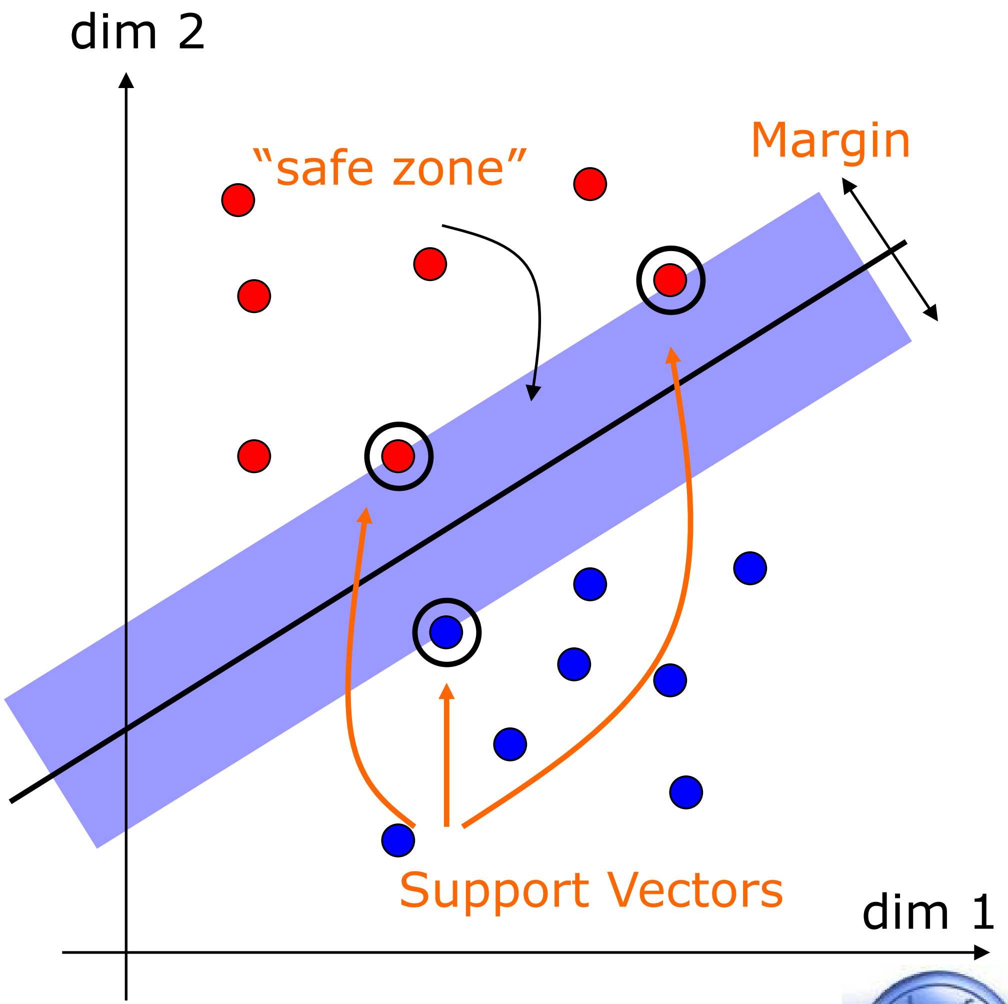
- How would you classify these points using a linear function in order to minimize the error rate?
- Linear function:
  - $f(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b$
  - A hyper-plane in the feature space
- Infinite number of answers!
- Which one is the best?



Linear SVM

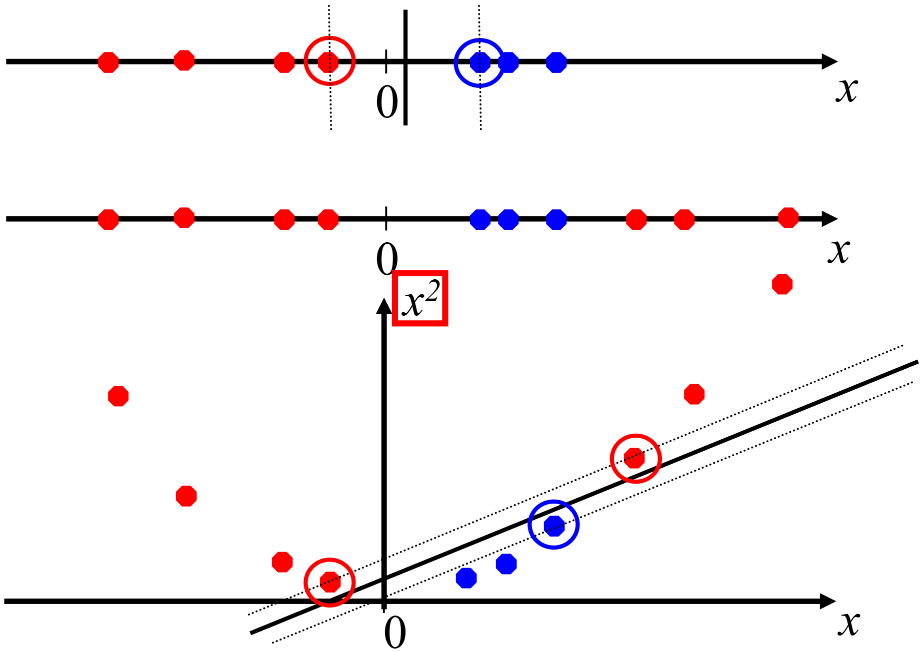
Large Margin Linear Classifier

- The linear function with the maximum margin is the best
- Margin is defined as the width that the boundary could be increased before hitting a data point
- Why it is the best?
  - Robust to outliers and thus strong generalization ability



Non-Linear SVM

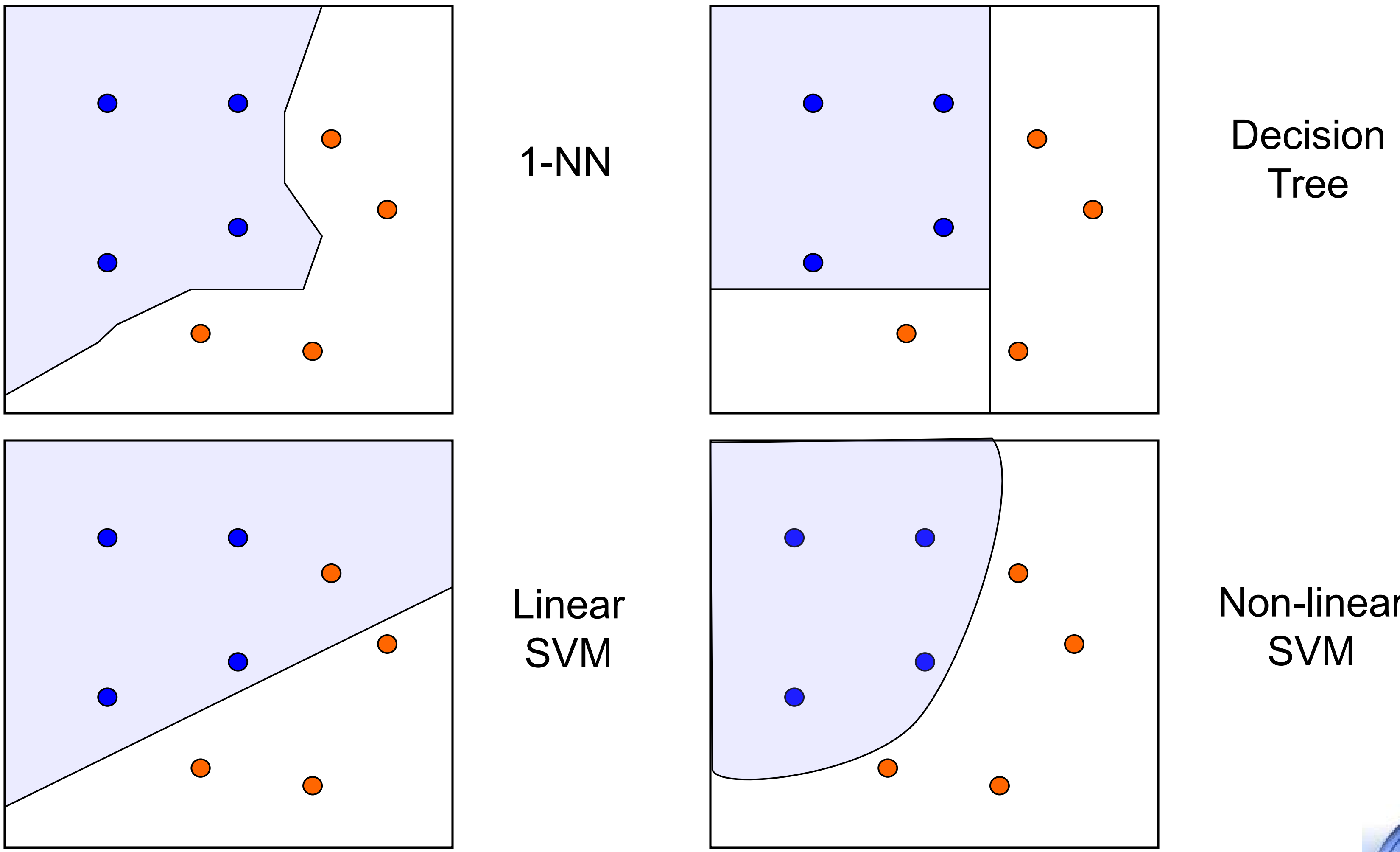
- Linearly separable:
- What if the data is too hard?
- How about…
  - mapping data to a high dimensional space



SVM
特点

- 在解决小样本、非线性问题上，具有独到的优势；因为对预测性能起关键因素的，是少数边界处的向量（支持向量）；只要边界处的向量分布正确、合理，预测效果就会较好
- 合适的变换核是关键，但对不同的问题，什么样的变换核有效，恰恰是个难点
- LibSVM：专门的平台

k-NN, Decision Tree and SVM



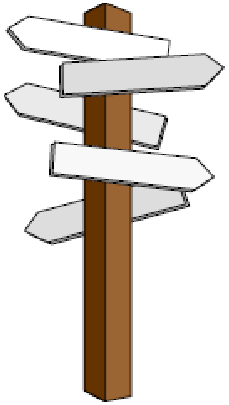
Other Learners

- Naïve Bayesian（朴素贝叶斯）
- Neural Networks（神经网络）
- Least Squares（最小二乘）
- Gaussian Mixture Models
- Hidden Markov Models（隐马尔科夫模型，时序问题）
- Dynamic Bayesian Network（动态贝叶斯网络，时序问题）
- etc.

Different methods are suitable for different applications

Road Map

- Machine Learning
- Learners
- Learning Paradigms
- Resources
- Summary



Learning Paradigms（学习范式）

- **Ensemble learning（集成学习）**
- **Deep Learning（深度学习）**
- Semi-supervised learning（半监督学习）
- Cost-sensitive learning（代价敏感性学习）
- Class-imbalance learning（类别不平衡学习）
- Multi-label learning（多标记学习）
- Multi-instance learning（多示例学习）
- ……

Focus on ideas rather than detailed algorithms

Ensemble Learning（集成学习）

An example: The gender recognition task

To predict whether a student is a boy or a girl

Useful features may include:

Height

Weight

Width of shoulder

Ensemble Learning（集成学习）

- Suppose 3 learners are trained with some collected data:
  —— SVM, Neural Network(NN), Decision Tree(DT)
- For a test set of 100 students, the best accuracy of the three classifiers are all 90% (10 students are wrongly classified).
- If the classification results are as follows:

  If only No.1-10 students are wrongly classified by SVM

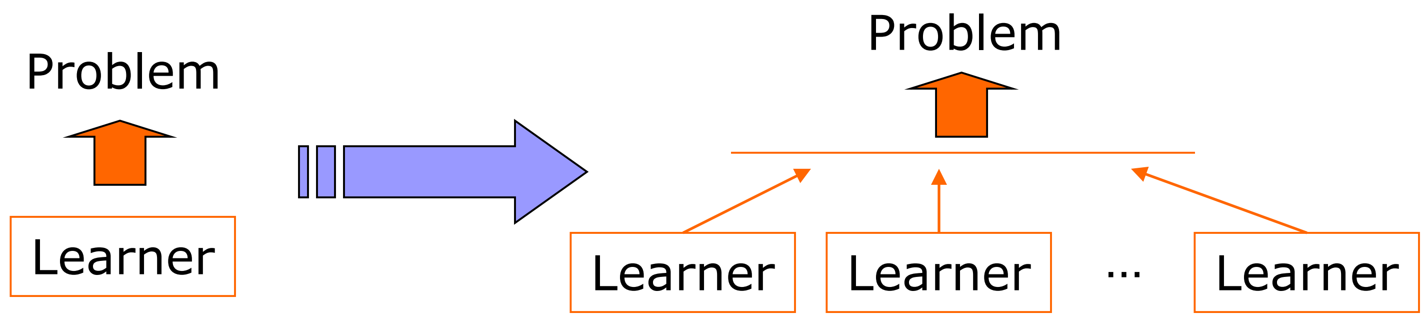
  If only No.11-20 students are wrongly classified by NN

  If only No.21-30 students are wrongly classified by DT

- Fuse 3 classifiers by majority voting, we can get 100% accuracy!

Ensemble Learning（集成学习）

- Training multiple individual learners for the same problem

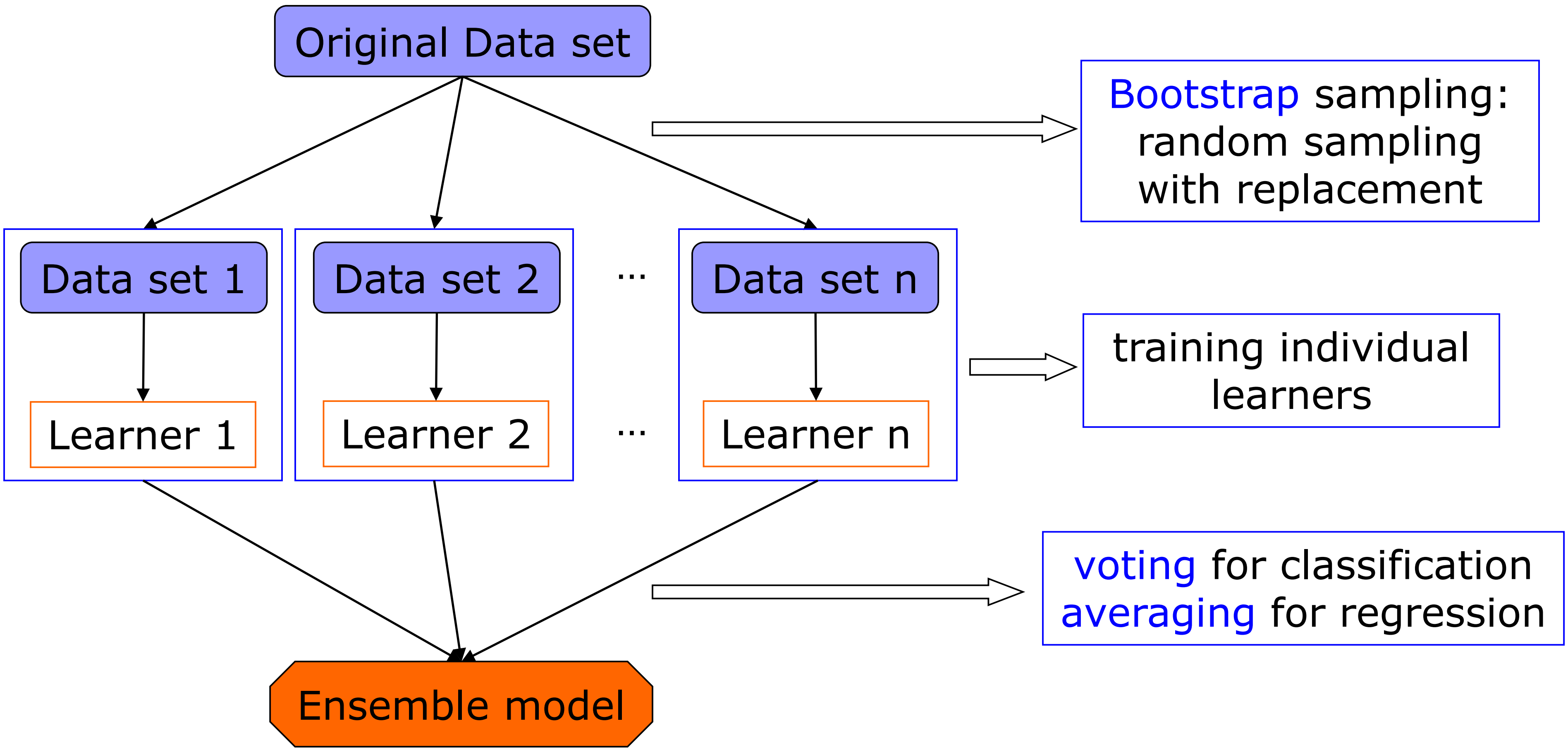


- Why ensemble?
  - The generalization ability of an ensemble is usually significantly better than the corresponding single learner

Base learners: the more accurate and the more diverse, the better

- Two steps:
  - (1) train base learners
  - (2) combine the individual predictions

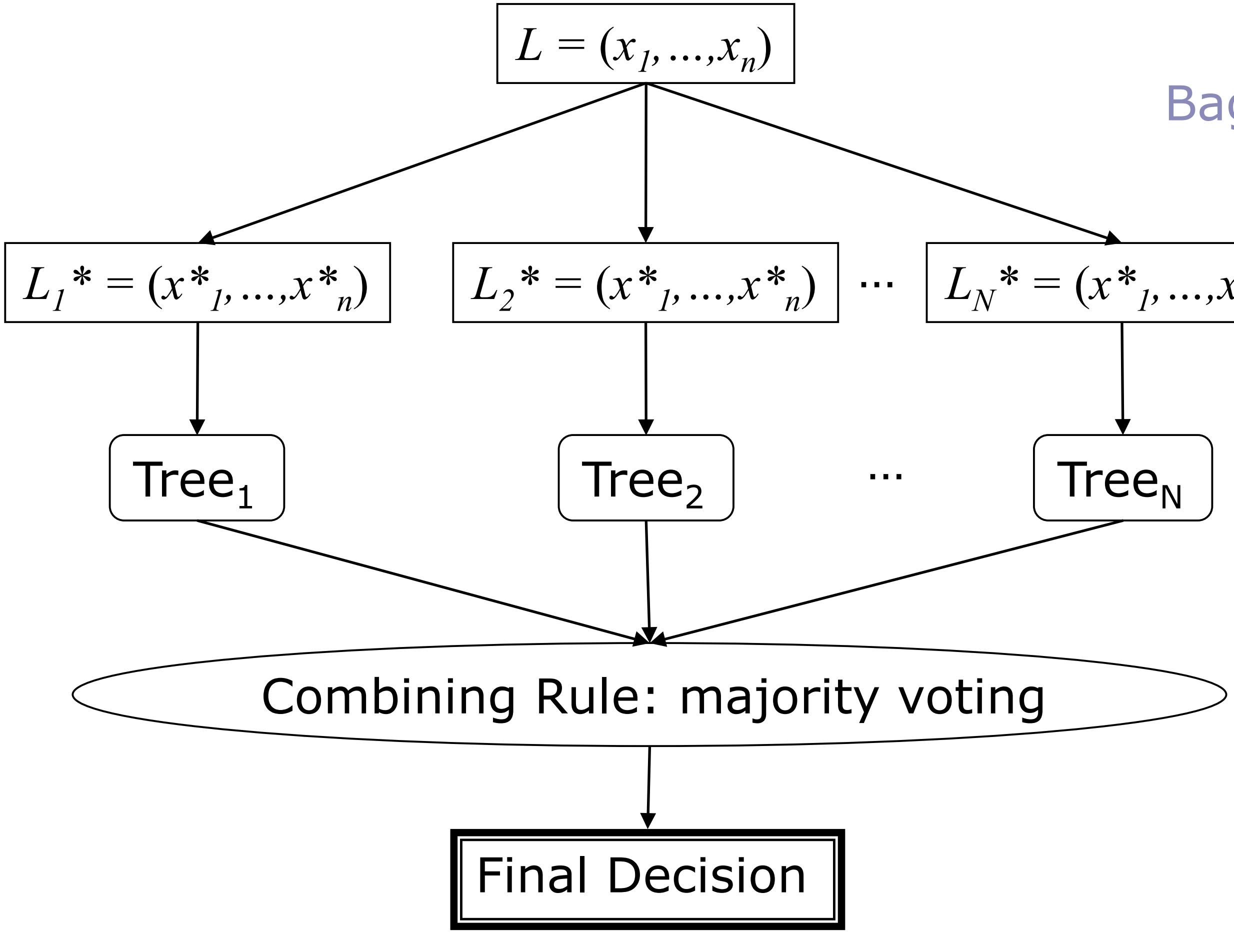
Bagging(并行)



Learner → decision tree: Random Forest **[Breiman, MLJ01]**

Selective ensemble: Many Could be Better Than All **[Zhou et al, AIJ02]**

Bagging with decision trees: Random Forest



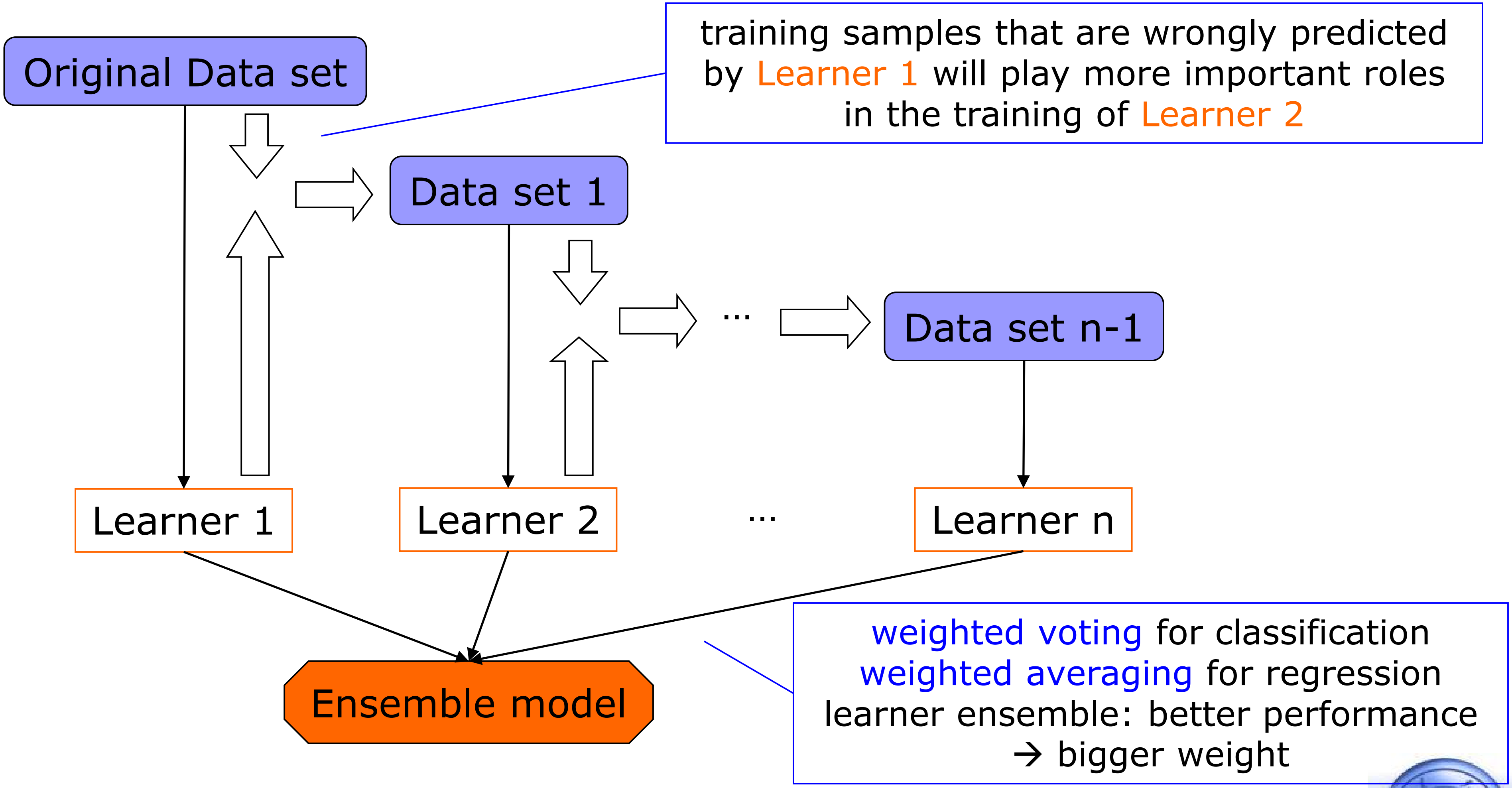
Bagging with decision trees

Tree growing process

1) no pruning;
2) m variables are used;

$m << M$: the only adjustable parameter to which random forests is somewhat sensitive

Boosting（串行）


training samples that are wrongly predicted by Learner 1 will play more important roles in the training of Learner 2
Original Data set
Data set 1
...
Data set n-1
Learner 1
Learner 2
...
Learner n
Ensemble model
weighted voting for classification
weighted averaging for regression
learner ensemble: better performance → bigger weight

Ensemble Learning（集成学习）

- Ensemble learning is not always helpful

Back to the gender recognition task(to 100 students):

If the three learners make the same mistakes（犯错同样的错误）

No.1-10 students are wrongly classified by SVM

No.1-10 students are wrongly classified by NN

No.1-10 students are wrongly classified by DT

Or if the performance of all three learners are poor（准确率都很低）

No.1-60 students are wrongly classified by SVM

No.11-70 students are wrongly classified by NN

No.21-80 students are wrongly classified by DT

- No improvement can be made by fusing the three learners.

Ensemble Learning（集成学习）

- The combined learners (classifiers) should have

1. relatively high accuracy (at least 51%)
2. high diversity(差异)

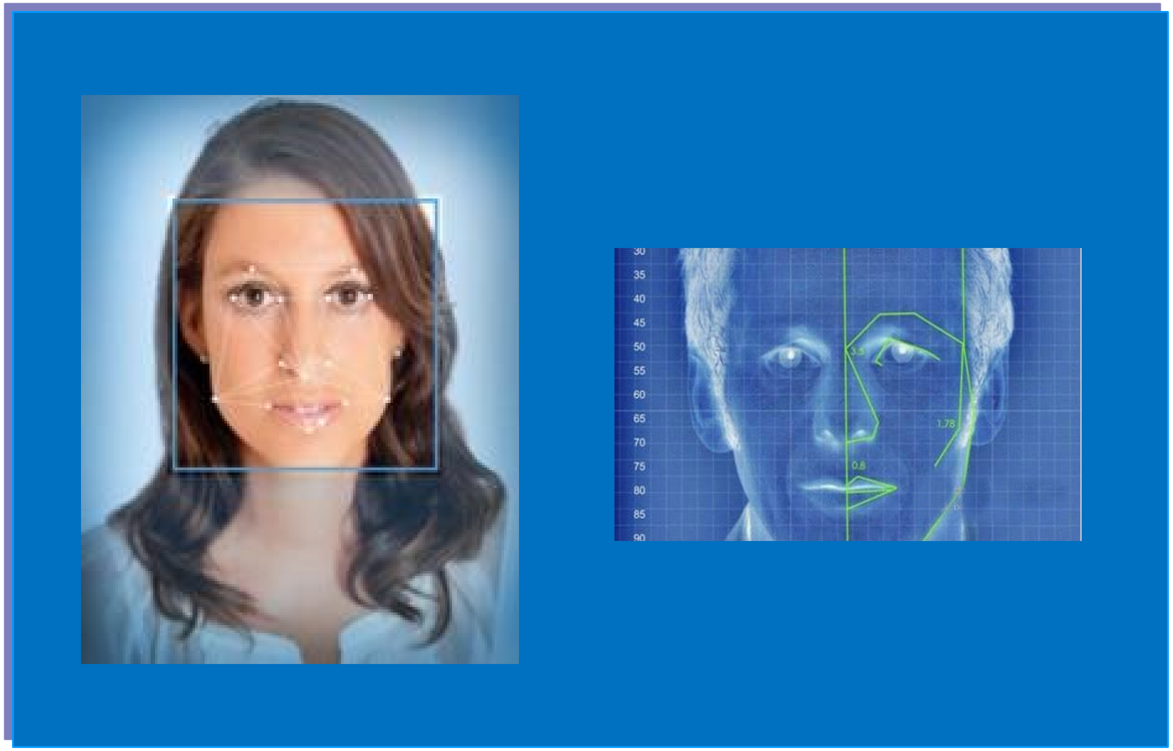
How to get "good and different" individual learners is the key point of ensemble learning.

Ensemble Learning（集成学习）：更为广义的理解

- 上面介绍的，是多学习器的集成，即，多个分类方法的集成
- 实际上，集成学习的思想，更为广义
- 还可以有更多不同水平的集成学习
  - Sensor level
  - Feature level
  - Score level
  - Decision level
  - … …
- 主要是看具体在哪个阶段进行集成

Ensemble Learning:
at sensor level

- 通过不同的设备，能够获取人脸的可见光图像和红外图像（不同的采集设备、不同的传感器）
- 同时使用一个人的可见光图像和红外图像，实现人脸识别



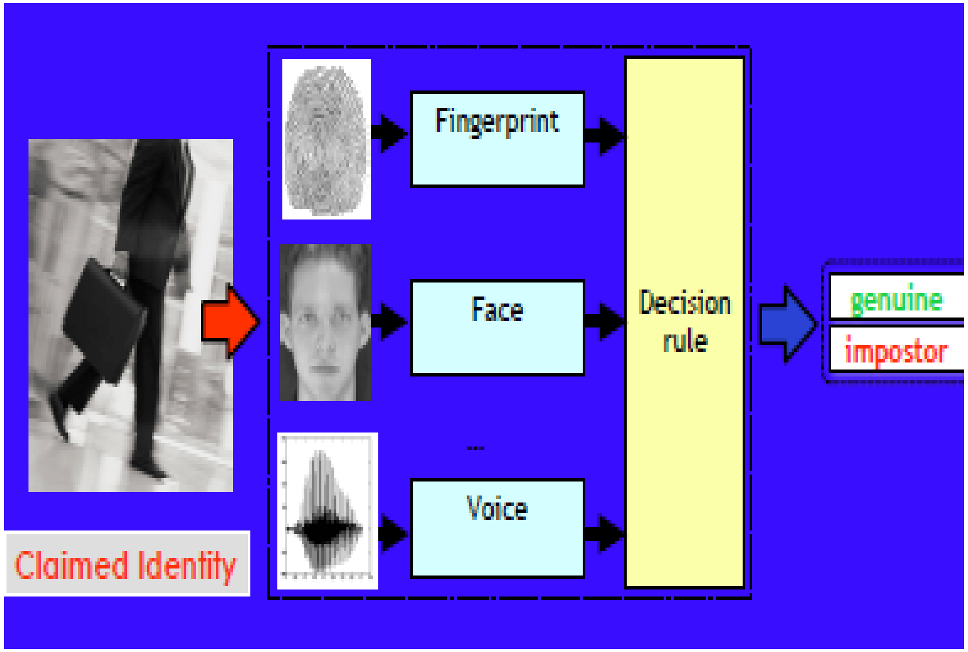
Ensemble Learning:
at feature level(multi-view)

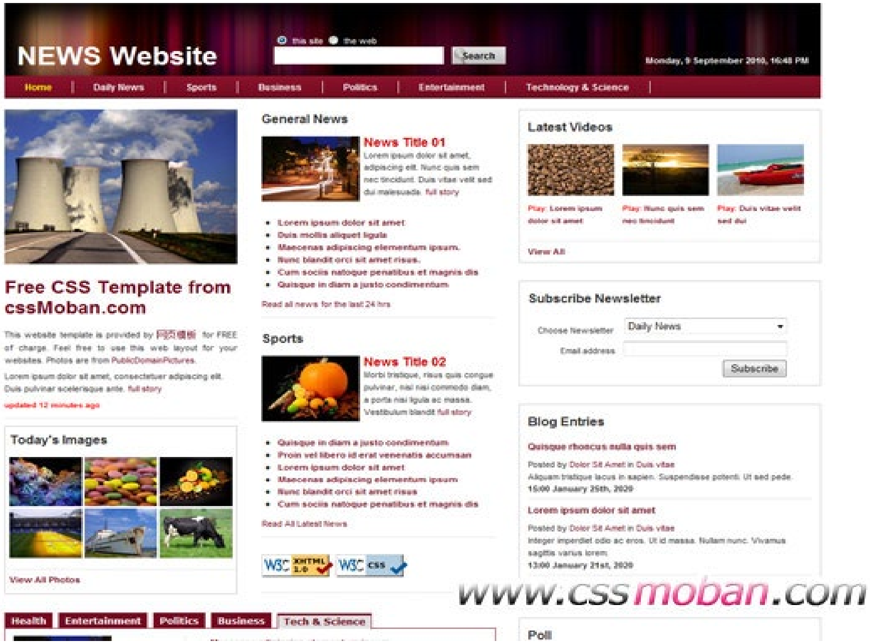
- Multimodal biometrics

  Fingerprint, face and voice are three views of a person

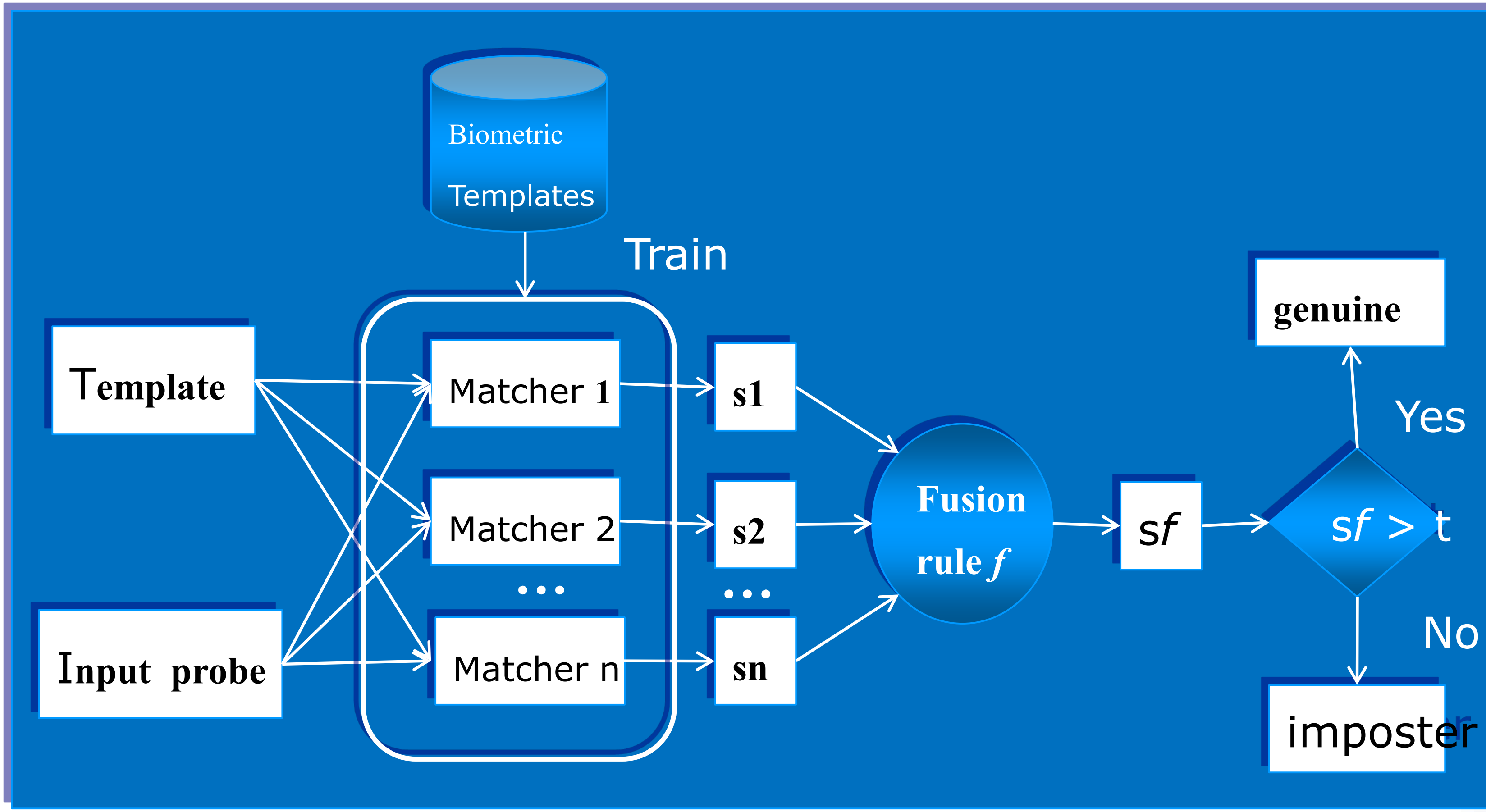
- Classification of Webpage

  Texts, images and hyper-links are three views of a webpage

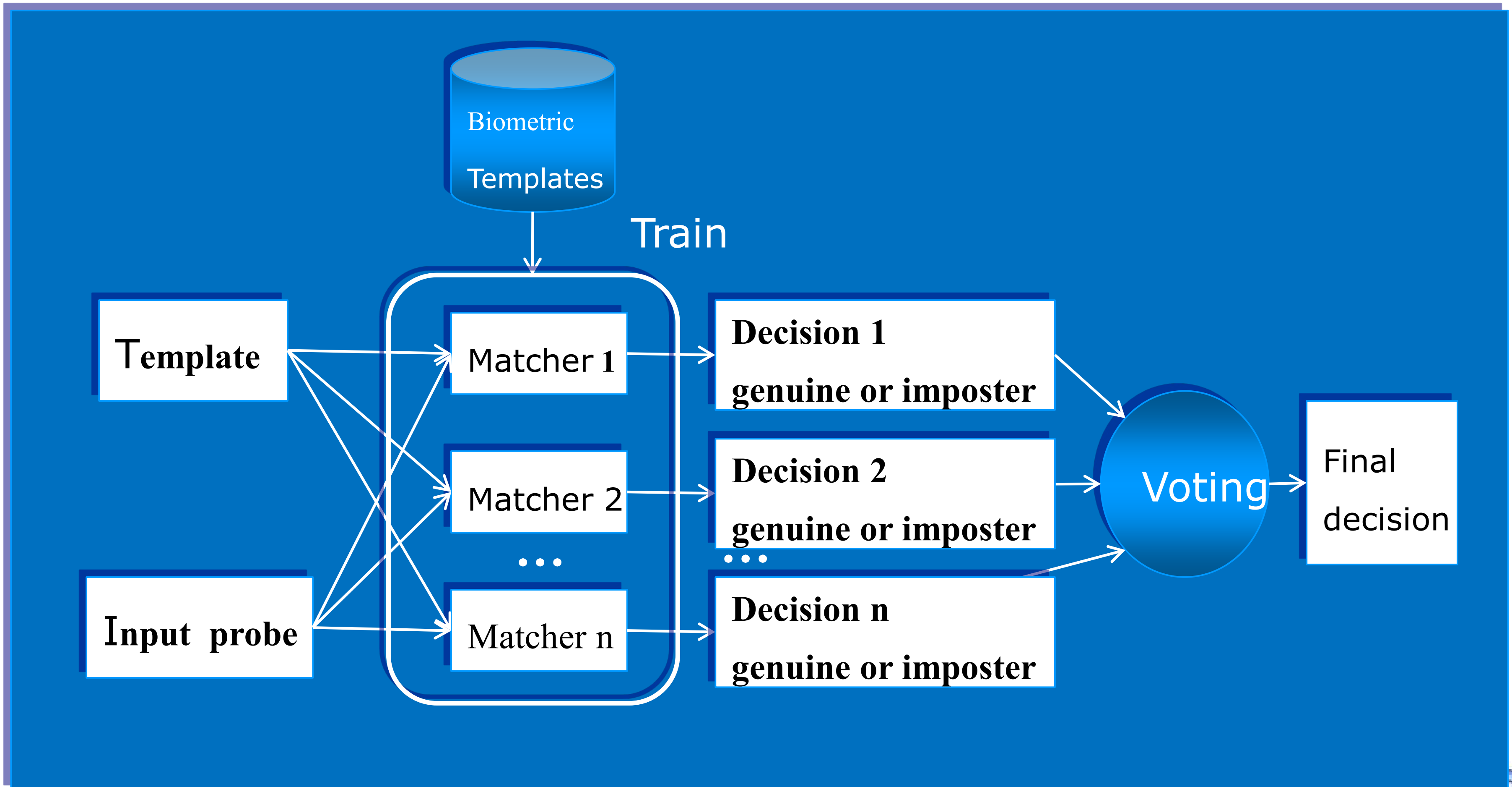




Ensemble Learning:
at score level



Ensemble Learning:
at decision level



Ensemble Learning（集成学习）

- 思想很好理解（以多取胜、资源代价换取性能提升）
- （集成）程序很好实现
- 可用于各种预测问题：方法论、独立于具体问题
- 几乎适用于所有领域：万金油
- 从集成学习范式提出到深度学习出现（2006）之前，曾很长时间在机器学习领域独领风骚
- 效果如何，主要看两个方面：（1）被集成的基学习器性能如何；（2）被集成的多个学习器是否存在差异性（互补性）
- 对数学建模而言，可考虑预先准备多种学习器（决策树、支持向量机、神经网络等）的程序，根据建模任务的需要，现场集成（可尝试不同的集成策略）

集成学习方法的巨大成功

- KDDCup'07: 1st place for "... Decision Forests and ..."
- KDDCup'08: 1st place of Challenge1 for a method using Bagging; 1st place of Challenge2 for "... Using an Ensemble Method "
- KDDCup'09: 1st place of Fast Track for "Ensemble ... "; 2nd place of Fast Track for "... bagging ... boosting tree models ...", 1st place of Slow Track for "Boosting ... "; 2nd place of Slow Track for "Stochastic Gradient Boosting"
- KDDCup'10: 1st place for "... Classifier ensembling"; 2nd place for "... Gradient Boosting machines ... "

集成学习方法的巨大成功

- KDDCup'11: 1st place of Track 1 for "A Linear Ensemble … "; 2nd place of Track 1 for "Collaborative filtering Ensemble", 1st place of Track 2 for "Ensemble …"; 2nd place of Track 2 for "Linear combination of …"
- KDDCup'12: 1st place of Track 1 for "Combining… Additive Forest…"; 1st place of Track 2 for "A Two-stage Ensemble of…"
- KDDCup'13: 1st place of Track 1 for "Weighted Average Ensemble" ; 2nd place of Track 1 for "Gradient Boosting Machine"; 1st place of Track 2 for "Ensemble the Predictions"

集成学习方法的巨大成功

- KDDCup'14: 1st place for "ensemble of GBM, ExtraTrees, Random Forest..." and "the weighted average" ; 2nd place for "use both R and Python GBMs"; 3rd place for "gradient boosting machines... random forests" and "the weighted average of..."
- KDDCup'15: 1st place for "Three-Stage Ensemble and Feature Engineering for MOOC Dropout Prediction"
- Netflix Prize:
  - ✓ 2007 Progress Prize Winner: Ensemble
  - ✓ 2008 Progress Prize Winner: Ensemble
  - ✓ 2009 $1 Million Grand Prize Winner: Ensemble !!

Ensemble Learning:
Tutorials

- L Rokach, Pattern Classification Using Ensemble Methods, World Scientific, Singapore, 2010
- Z-H Zhou, Ensemble Methods: Foundations and Algorithms, Chapman & Hall/CRC, Boca Raton, 2012
- Z-H Zhou, Machine Learning, Chapter 8, 2016

Ensemble Learning:
Useful web pages and tools

- Web pages:
  - http://www.scholarpedia.org/article/Ensemble_learning
  - https://en.wikipedia.org/wiki/Ensemble_learning
  - https://beta.learning.intersystems.com/course/view.php?id=15
- Online Tools
  - The Waffles (machine learning) toolkit

  https://en.wikipedia.org/wiki/Waffles_(machine_learning)

Deep Learning（深度学习）

当前最火的机器学习范式，在图像分类、物体检测与识别、语音识别等领域，取得了突破性进展

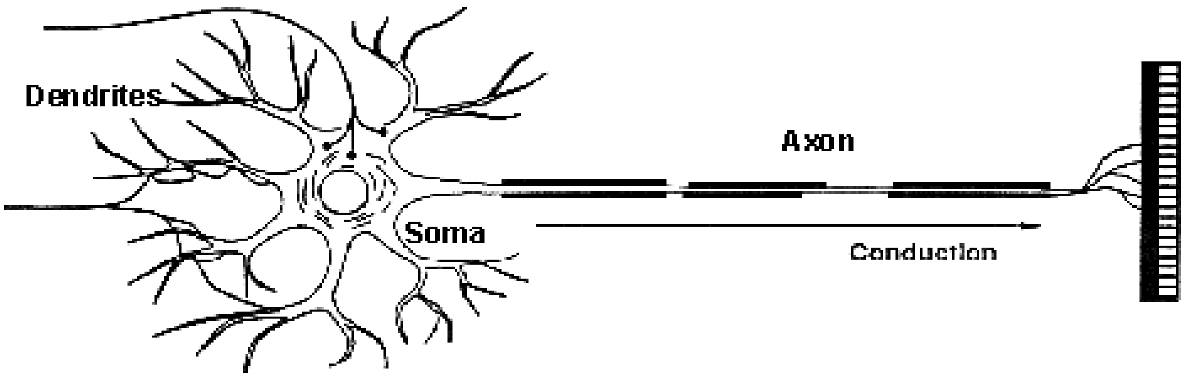
- 人工神经网络
- 多隐层人工神经网络
- 什么是深度学习
- 深度学习的发展历史
- 取得的突破性进展
- 为何有效
- 什么情况下应该考虑使用
- 发展趋势和局限
- 有用资源

Artificial Neural Network (ANN)

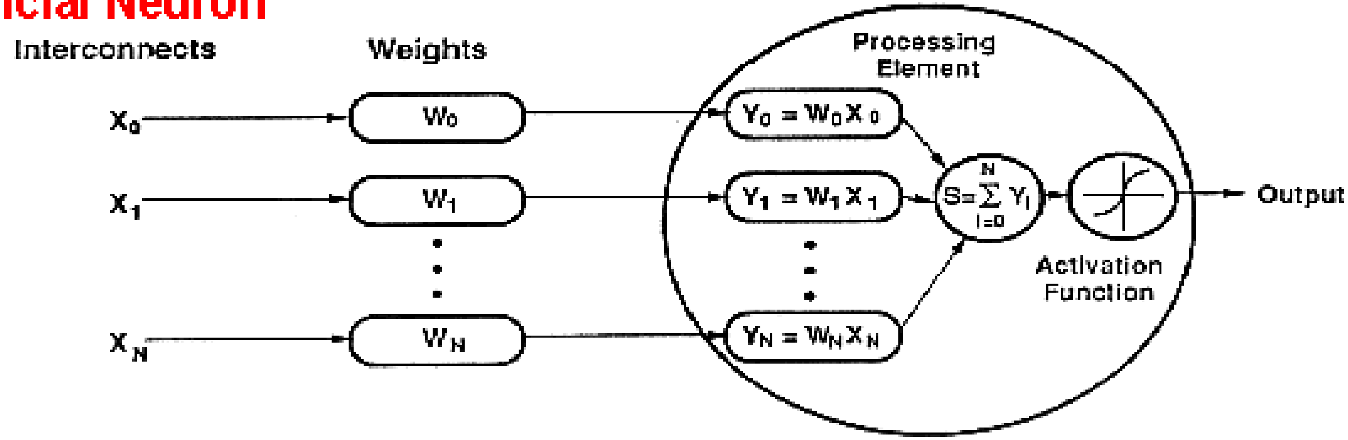
Our brains are a huge network of processing elements. A typical brain contains a network of 10 billion neurons.

Biological Neuron and ANN

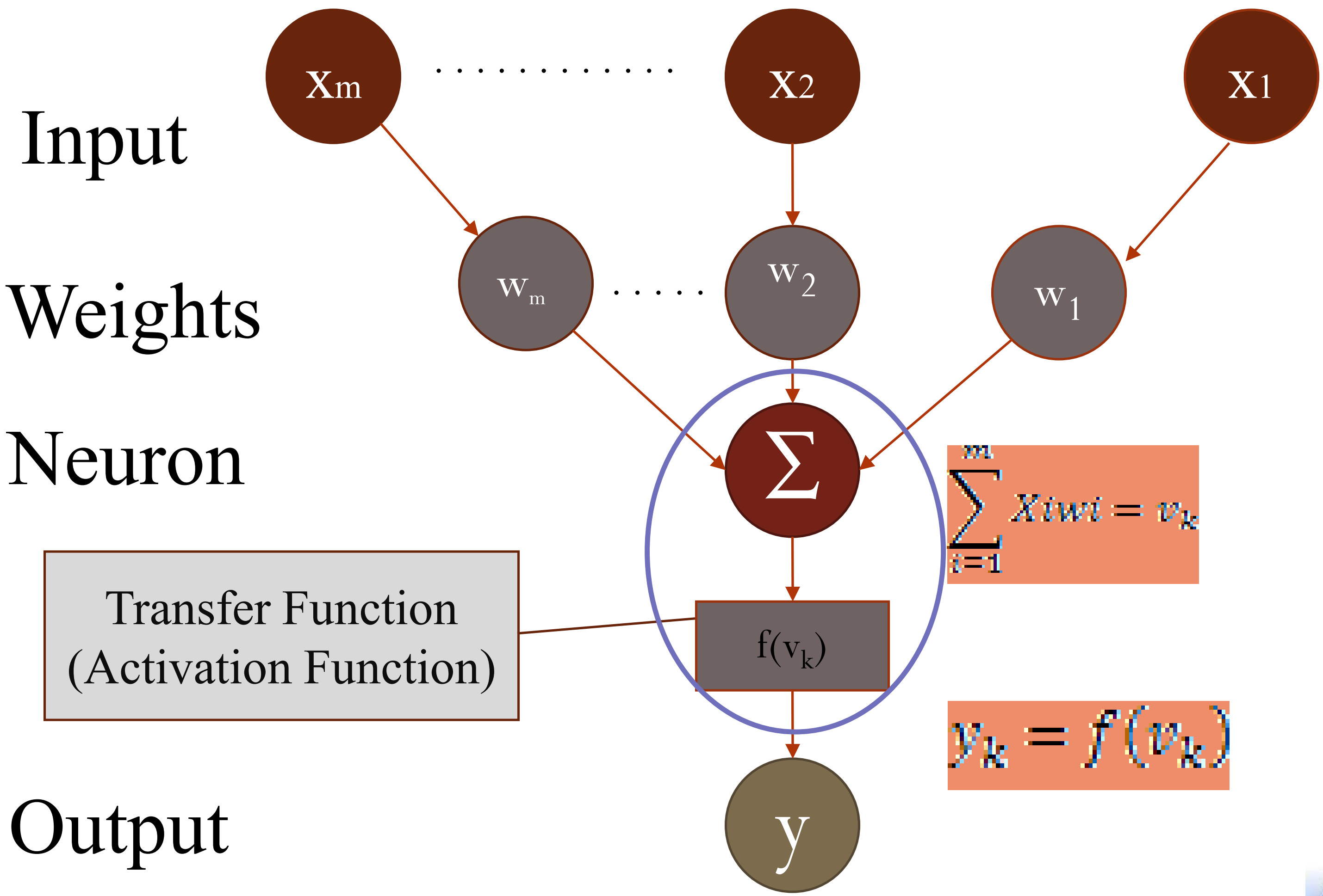
Biological Neuron



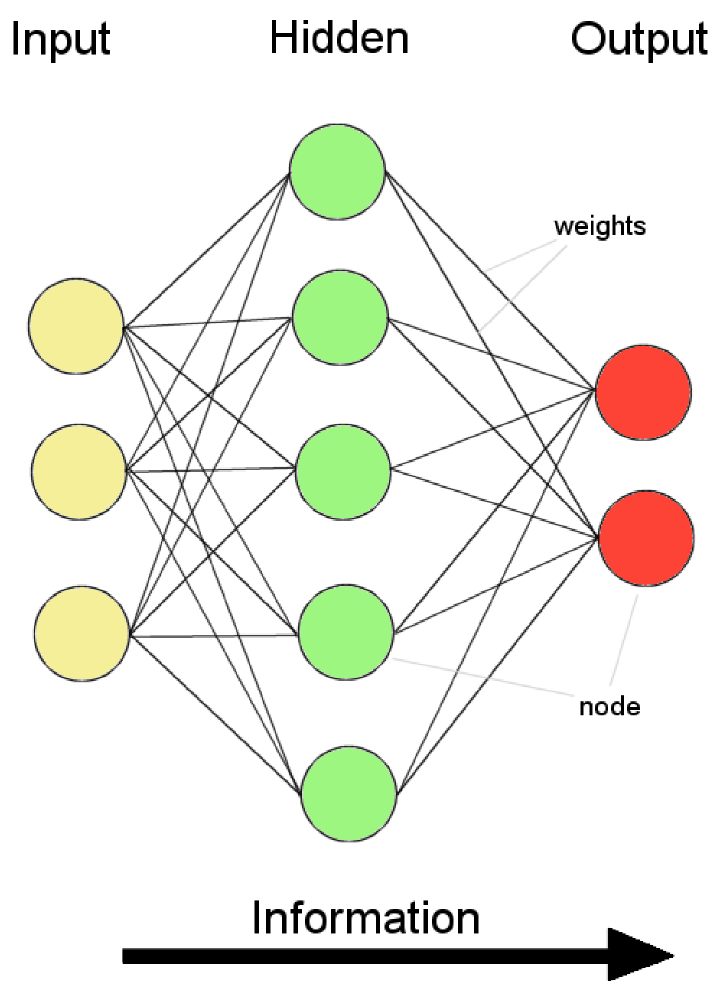
Artificial Neuron



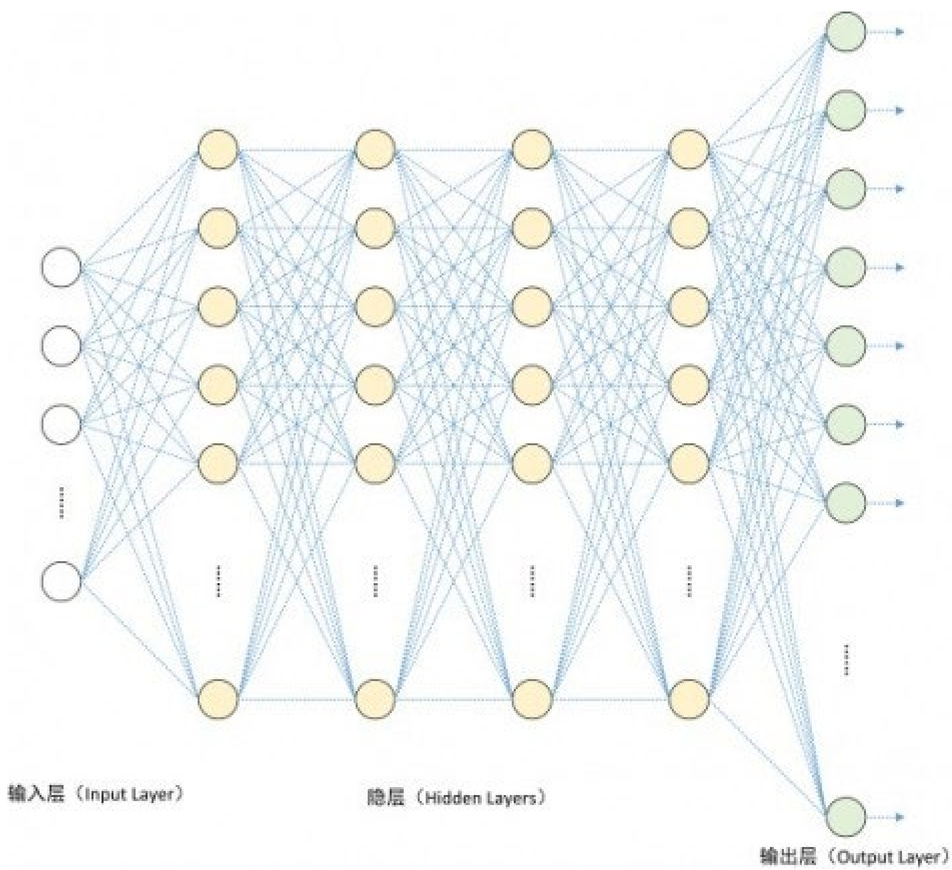
ANN--单层感知器


Input
Weights
Neuron
Output
Xm
X2
X1
wm
w2
w1
Σ
Transfer Function
(Activation Function)
f(vk)
y
$\sum_{i=1}^{m} Xiwi = v_k$
$y_k = f(v_k)$

ANN—单隐层神经网络



ANN—多隐层神经网络



Deep Learning（深度学习）：What is it

- 本质上，就是多隐层人工神经网络
- 属于深层模型（Deep Model），SVM等常见学习器都属于浅层模型（Shallow Model）
- 针对具体任务，利用给定的一批标记数据，先训练一个多隐层神经网络，然后使用它，这就是深度学习
- 深度学习与浅层模型几个突出的不同点

1. 自动学习特征 Vs. 经验知识+人工定义特征
2. 端到端（end to end）Vs. 分步、分治
3. 超强的非线性建模能力 Vs. 有限的非线性建模能力

Deep Learning（深度学习）：History

- 与多隐层神经网络有关的几个重要时间节点
  - 上世纪80年代就出现了多隐层人工神经网络
  - Before 2000（受到的关注少）
  - 2000-2005（Bengio等，在推动深度学习，但在应用上取得的进展有限）
  - 2006至今（Hinton等在《科学》发表论文：优异特征学习能力、无监督的分层预训练）：进入深度学习的时代
  - 优化方法、优化技巧在近几年深度学习的发展中起了很重要的作用

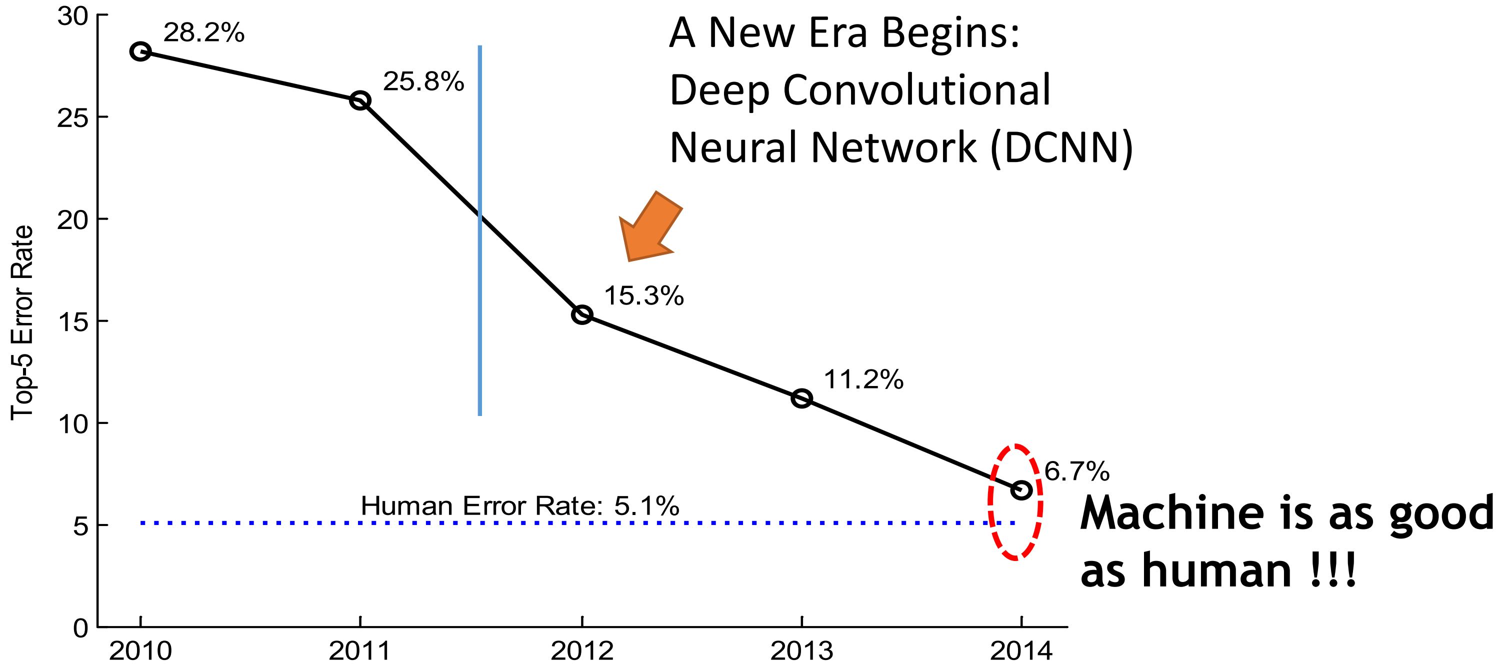
Deep Learning（深度学习）：Why it could not work well before 2006

- 设备条件限制：缺乏高性能计算装备，更没有GPU
- 数据条件限制：缺乏大数据（训练数据）
- 技术条件限制：对多隐层神经网络，缺乏有效的训练方法
- 2006年以后，上述限制逐步解除了

突破性进展1：
图像分类：on ImageNet

- Database:
  1000 categories,
  1.2 million training images,
  150,000 testing images.
- Task:
  classify testing image into one of 1000 categories.

突破性进展1：
图像分类：on ImageNet



突破性进展2：人脸识别：on LFW

- LFW是用于人脸识别的一个相对高难度、大尺寸公开数据库
- 基于深度学习的DeepID人脸识别技术在LFW库上的准确率达到99.77%，比人眼识别更加精准!

突破性进展3：
深度学习在中文语音识别上超过了人类

- 2015年12月，百度研究院硅谷人工智能实验室
- 中文语音测试：人类语音识别的错误率是4.0%，而机器是3.7%
- 百度首席科学家吴恩达:对于无上下文的短语，基于深度学习的计算机系统的识别能力超过了人类

突破性进展4：
围棋人机大战：AlphaGo Vs. 李世石

- 在上世纪90年代末期以前，普遍认为：尽管机器可在国际象棋比赛中战胜人类棋手，但机器永远不可能，至少在可以预见的很长时间内，机器不可能在围棋上战胜人类
- 2016年3月9-15日，AlphaGo以4:1战胜韩国九段棋手李世石
- 当前世界围棋界，应该没有人可以战胜AlphaGo了，今后，机器的优势将更加突出
- 深度学习，是AlphaGo使用的核心技术之一

Deep Learning（深度学习）：最新进展

- Attention（注意力机制）
- Reasoning（推理）
- Planning and Reinforcement Learning （规划与强化学习）
- 当前，最前沿的是融合技术。比如，视觉与自然语言理解的结合
- 向后看，可能的前沿方向：无人驾驶、推理和回答问题

Deep Learning（深度学习）：为何展示出极其突出的性能优势？

- 具有自动学习特征的能力（Feature Learning）
- 学习到的特征体系和人工定义特征不同，在完备性和非冗余性上，更准确地说，在区分性上，强于人工定义的特征
- 对复杂分类问题，有能力学习到极其复杂的“分界面”（过拟合，非贬义）
- 解决问题的思路和技术框架，都具有较强的通用性

Deep Learning（深度学习）：什么情况下应该考虑使用深度学习

- 针对任务要求，难以人工定义特征（因为这需要先验知识），或者人工定义的特征不够有效：Feature Learning
- 大量标记样本（期望训练集分布更接近全集分布）。但对标记样本数量的要求也不是那么绝对（很多时候，使用少量标记样本也很有效，核心还是分布问题）
- 高性能计算资源：GPU（尤其是对图像、语音）；但这是针对训练过程；应用时，普通电脑就够了
- 有效的网络训练方法（这是对科研而言，对竞赛，使用已有的、相对成熟的网络训练方法即可）
- 面向数学建模：（1）使用开源框架和相近数据，预先训练网络；（2）竞赛现场，利用给定数据（数据量一般不大），进行再训练（训练量应该不大），这是迁移学习的概念

Deep Learning：有用资源

▶http://blog.csdn.net/zouxy09/article/details/8775360

了解一些deep learning基本方法的思想

▶http://ufldl.stanford.edu/wiki/index.php/UFLDL教程

deep learning大牛Andrew Ng所写，还有实验、源代码，推荐细读

Deep Learning：有用资源

- Platforms:
  - Pytorch：https://pytorch.org/
  - TensorFlow: https://tensorflow.google.cn/
  - Keras：https://keras.io/

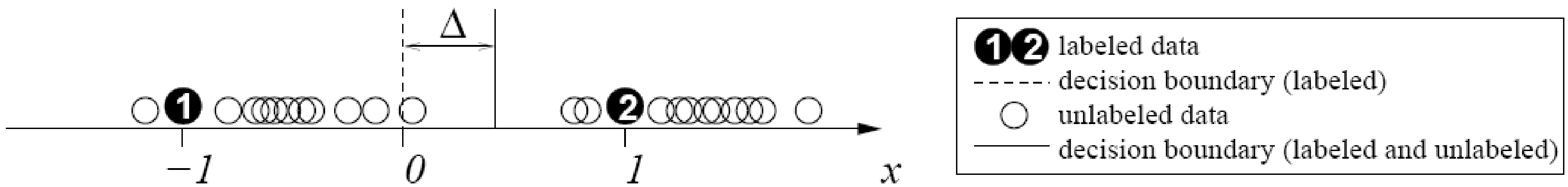
Semi-Supervised Learning

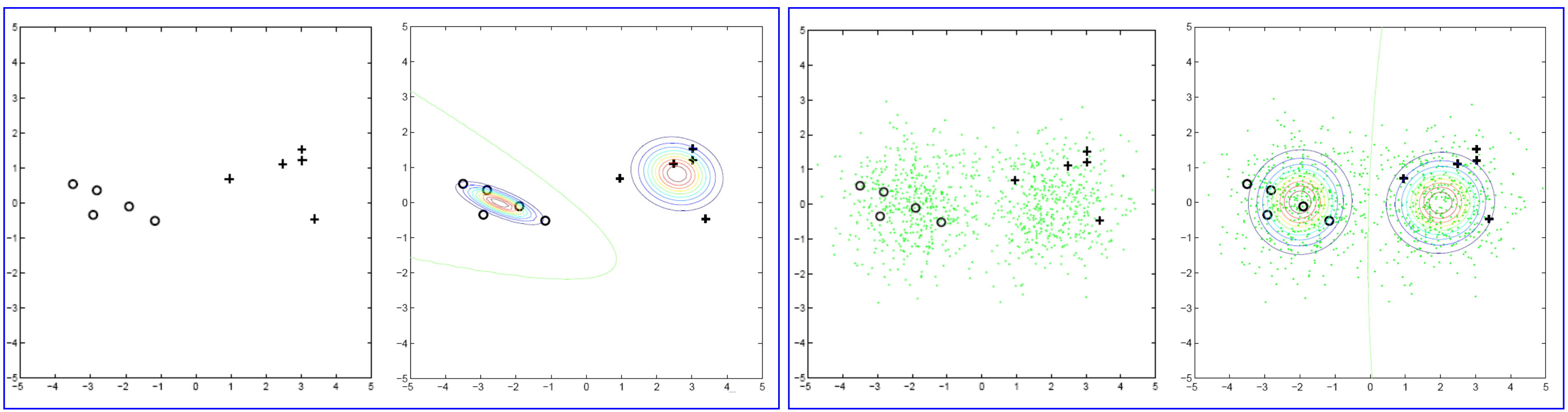
- Why semi-supervised?
  - $\mathbf{X}_{1:n}$ known, $\mathbf{Y}_{1:l}$ known, $\mathbf{Y}_{l+1:n}$ unknown
  - labeled data may be hard to get; unlabeled data is cheap
  - people want better performance for free

- Example: Web Page Classification

Learnability

- Goal: using both labeled and unlabeled data to build better learners, rather than using each one alone.





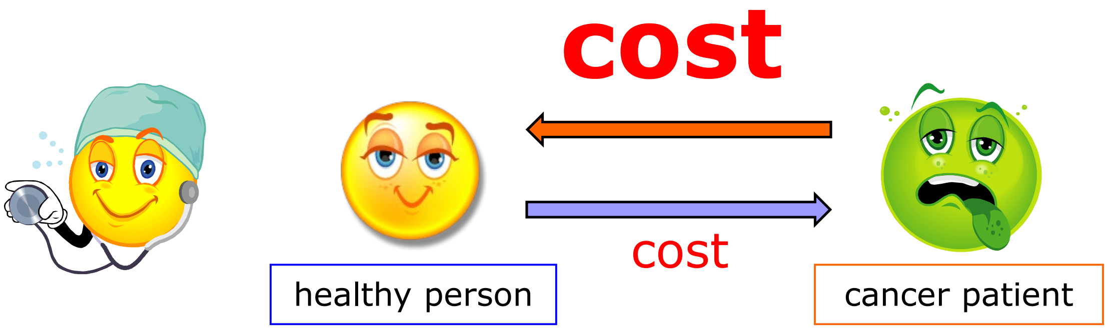
labeled data　　GMM model　　adding unlabeled data　　GMM model

Semi-Supervised Learning Algorithms

- Self-training
- Generative models
  - Gaussian mixture models + EM
- Semi-Supervised SVM (S3VM)
  - a.k.a.: Transductive SVM (TSVM)
- Disagreement-based algorithms
  - Co-training
  - Tri-training
- Graph-based algorithms
  - Label propagation
  - Manifold regularization
  - local and global consistency

Cost-Sensitive Learning

- Why cost-sensitive learning?
  - Traditional view: low error rate → good performance
  - However, in many real applications, different mistakes often have different costs
  - We should minimize the total cost instead of simply minimizing the error rate



Keys: (1) estimate the misclassification costs;
(2) minimize the total cost;

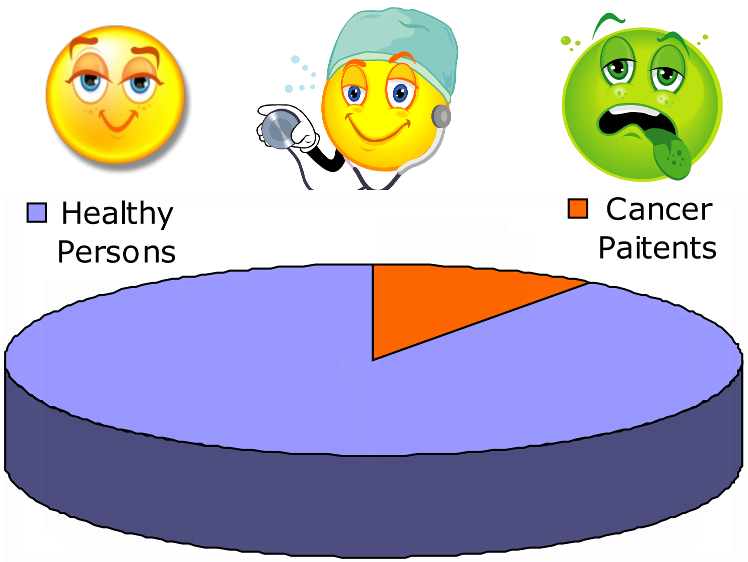
CSL Algorithms

- Direct modification on traditional learners
  - cost-sensitive SVM
  - cost-sensitive decision tree
  - cost-sensitive neural networks
  - cost-sensitive boosting

- Rescaling
  - re-weighting
  - re-sampling
    - undersampling & oversampling
  - threshold moving
    - MetaCost

Class-Imbalance Learning

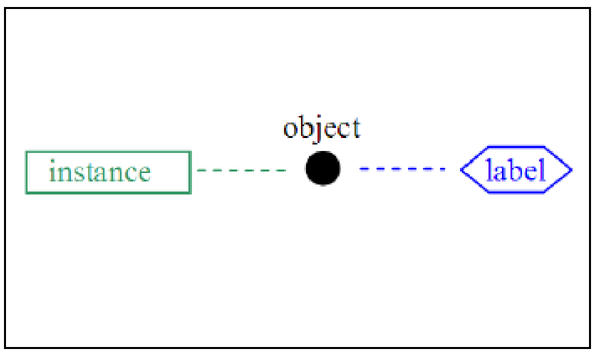
- Why class-imbalance learning?
  - In many real applications, the data sets are typically imbalanced, i.e., some classes have much more instances than others.
- Cancer detection
  - healthy : cancer = 99:1
  - minimizing error rate: classify all instances as healthy → 1% error rate



- CIL algorithms
  - cost-sensitive learning: re-sampling, re-weighting, etc.
  - one-class learning: one-class SVM

Multi-Label Learning

- Why multi-label?
  - In traditional supervised learning: a real-world object is represented by an instance
  - The instance is associated with a label which indicates the concerned characteristics of the object



but

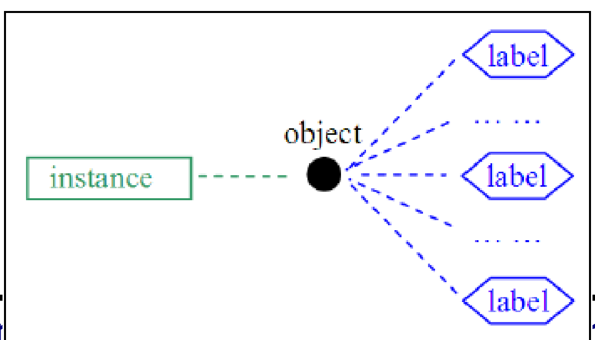
Elephant?

Lion?

Bush(丛林)?

Tropic(热带)?

Africa?



Multi-label learning: an object is attached with multiple labels

MLL Algorithms

- Decomposing the task into multiple binary classification problems each for a class
  - MLSVM
- Considering the ranking among labels
  - BoosTexter
  - BP-MLL
  - RankSVM
- Exploring the label correlation
  - Probabilistic generative models
  - Maximum entropy methods

Elephant?

Lion?

Bush?

rank1

Tropic?

Africa?

rank2

Penguin(企鹅)?

Iceberg(冰山)?

Multi-Instance Learning

- Why multi-instance?
  - Multi-label learning only addresses the output ambiguity
  - How about the input ambiguity?

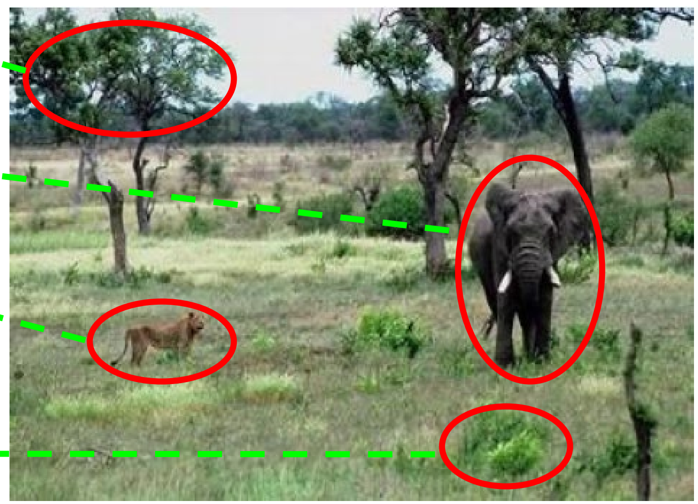
tree
$[a_1,a_2,\ldots,a_d]$

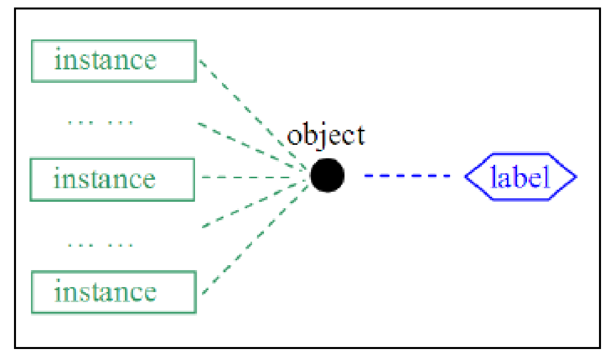
elephant
$[b_1,b_2,\ldots,b_d]$

lion
$[c_1,c_2,\ldots,c_d]$

……

grass
$[u_1,u_2,\ldots,u_d]$





total image ⇨ a bag

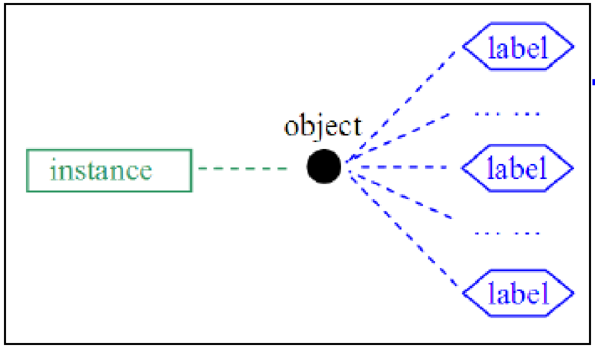
regions in image ⇨ instances in bag

Multi-instance learning:
an object contains multiple instances

MIL Algorithms

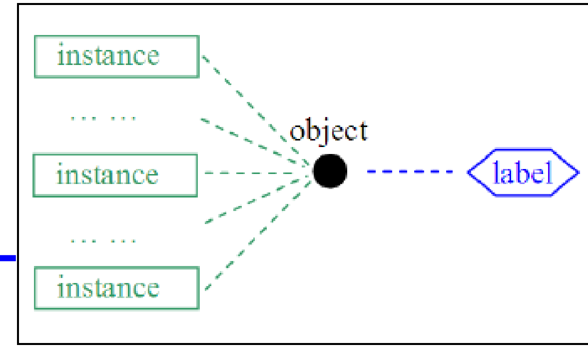
- Direct modification on traditional learners
  - by shifting their focuses from the discrimination on the instances to the discrimination on the bags
  - k-NN → Bayesian-kNN, Citation-kNN
  - decision tree → Relic, ID3-MI, RIPPER-MI
  - SVM → MI-SVM, mi-SVM, DD-SVM
- Other topics
  - density estimation: Diverse Density, EM-DD
  - kernel computation: multi-instance kernels
  - regression: MI-LR
  - clustering: BAMIC
  - ensemble: MI-Ensemble, MI-Boosting

Multi-Instance Multi-Label Learning



MLL considers output ambiguity

MIL considers input ambiguity



Input and output ambiguities usually occur simultaneously!

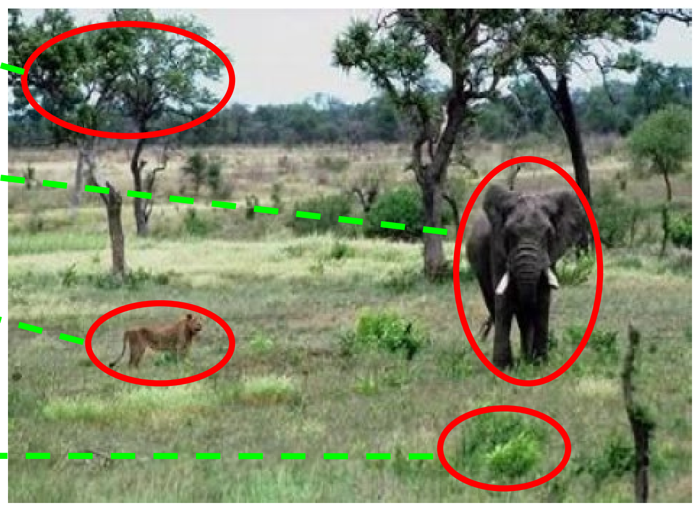
tree
$[a_1, a_2, \ldots, a_d]$

elephant
$[b_1, b_2, \ldots, b_d]$

lion
$[c_1, c_2, \ldots, c_d]$

......

grass
$[u_1, u_2, \ldots, u_d]$



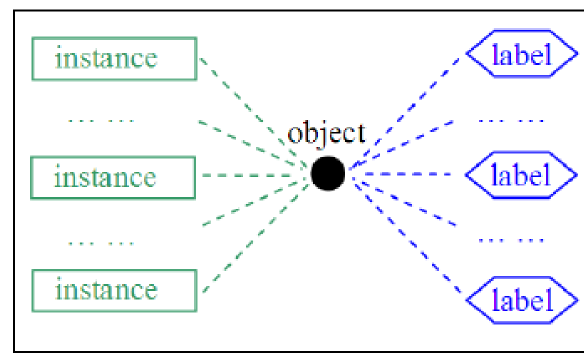
Elephant?

Lion?

Bush?

Tropic?

Africa?



Multi-instance multi-label learning (MIML):
each object contains many instances and is attached with multiple labels

Other Learning Paradigms

- Learning to Rank
  [Liu, FTIR09] & Hang Li

- Online Learning & Incremental Learning
  [Shwartz, Thesis07] & Yoram Singer

- Transfer Learning
  [Pan & Yang, TKDE (in press)] & Qiang Yang

Other Learning Paradigms

- Multi-Task Learning

  [Evgeniou et al, JMLR05] & [Argyriouet et al, NIPS’06] & Andreas Argyriou

- Reinforcement Learning

  [Kaelbling et al, JAIR96] & [Sutton & Barto, Book98]

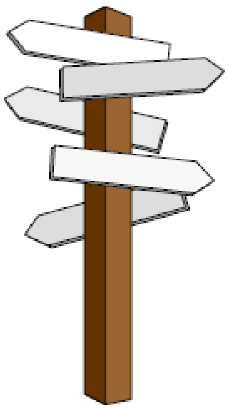
- Active Learning

  [Tong, Thesis01]

- etc.

Road Map

- Machine Learning
- Learners
- Learning Paradigms
- Resources
- Summary

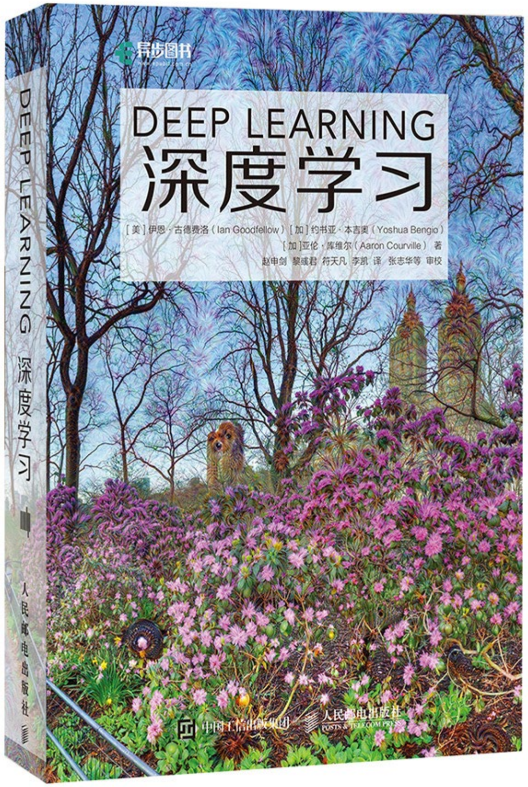


Resources: Books

- 周志华著，《机器学习》，中文， 清华大学出版社， ISBN号：978-7-302-42328-7，定价：88元，2016年1月出版（特别推荐）

Resources: Books

- Ian Goodfellow, Yoshua Bengio, Aaron Courville. Deep Learning, 2017.



Resources: Books

- R. Duda, P. Hart, D. Stork: Pattern Classification, 2nd Edition, Wiley, 2000（入门书）
- C. Bishop: Pattern Recognition and Machine Learning, Springer, 2006（难度大一些）
- T. Mitchell: Machine Learning, McGraw Hill, 1997（老一些）

Top Conferences/Top Journals

- Top Conferences
  - ICML, COLT, NIPS , ACML, etc. （侧重机器学习理论）
  - IJCAI, AAAI, ICCV, CVPR, ECCV, etc.（侧重机器学习应用）

- Top Journals
  - JMLR, AI,TPAMI, IJCV,TKDD, TKDE, PR, etc.

Resources: International Scholars

- **Tom Mitchell**: http://www.cs.cmu.edu/~tom/
- **Michael Jordan**: http://www.cs.berkeley.edu/~jordan/
- **Geoffrey Hinton**: http://www.cs.toronto.edu/~hinton/
- **Bernhard Schölkopf**: http://www.kyb.mpg.de/~bs/
- **Alexander Smola**: http://alex.smola.org/

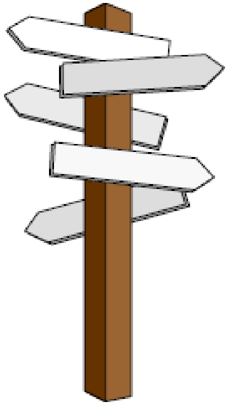
- **Rong Jin**: http://www.cse.msu.edu/~rongjin/
- **Jieping Ye**: http://www.public.asu.edu/~jye02/
- **Tong Zhang**: http://www.stat.rutgers.edu/~tzhang/
- **Andrew Ng**: http://ai.stanford.edu/~ang/
- **Eric Xing**: http://www.cs.cmu.edu/~epxing/
- **Fei Sha**: http://www-rcf.usc.edu/~feisha/
- **Xiaojin Zhu**: http://pages.cs.wisc.edu/~jerryzhu/

Resources: Tools

- Developing tools
  - MATLAB: http://www.mathworks.com/
  - WEKA: http://www.cs.waikato.ac.nz/~ml/weka/
  - LIBSVM: http://www.csie.ntu.edu.tw/~cjlin/libsvm/
  - MOSEK: http://www.mosek.com/
- Data sets
  - UCI ML Repository: http://archive.ics.uci.edu/ml/
  - UCI KDD Archive: http://kdd.ics.uci.edu/
  - Clustering data: http://cs.joensuu.fi/sipu/datasets/
- Google and WIKIPEDIA !

Road Map

- Machine Learning
- Learners
- Learning Paradigms
- Resources
- Summary



Summary

- Machine learning
  - concept
  - applications
- Learners
  - k-NN
  - decision tree
  - SVM
- Learning paradigms
  - ensemble learning
  - deep learning

References

- [Mitchell, Book97] T. Mitchell. Machine Learning. McGraw Hill, New York, 1997.
- [Quinlan, MLJ86] J. Quinlan. Induction of Decision Trees. Machine Learning, 1(1): 81-106, 1986.
- [Brieman et al, Book84] L. Breiman, J. Friedman, R. Olshen and C. Stone. Classification and Regression Trees. Chapman & Hall, New York, 1984.
- [Quinlan, Book93] J Quinlan. C4.5: Programs for Machine Learning. Morgan Kaufmann, San Francisco, CA, 1993.
- [Vapnik, Book95] V. Vapnik. The Nature of Statistical Learning Theory. Springer, New York, 1995.
- [Zhou, EB09] Z.-H. Zhou. Ensemble Learning. In: Encyclopedia of Biometrics. Springer, Berlin, 270-273, 2009.
- [Breiman, MLJ96] L. Breiman. Bagging Predictors. Machine Learning, 24(2): 123-140, 1996.
- [Breiman, MLJ01] L. Breiman. Random Forests. Machine Learning, 45(1): 5–32, 2001.

References

- [Zhou et al, AIJ02] Z.-H. Zhou, J. Wu and W. Tang. Ensembling Neural Networks: Many Could Be Better Than All. Artificial Intelligence 137(1-2): 239–263, 2002.
- [Freund & Schapire, JCSS97] Y. Freund and R. Schapire. A Decision-Theoretic Generalization of Online Learning and an Application to Boosting. Journal of Computer and System Sciences, 55(1): 119–139, 1997.
- [Zhu, TR08] X. Zhu. Semi-Supervised Learning Literature Survey. Computer Sciences TR 1530, University of Wisconsin Madison, 2008.
- [Ling & Sheng, EML08] C. Ling and V. Sheng. Cost-Sensitive Learning and the Class Imbalance Problem. In: Encyclopedia of Machine Learning. Springer, Berlin, 2008.
- [Weiss, KDDEXP04] G. Weiss. Mining With Rarity: A Unifying Framework. SIGKDD Explorations, 6(1): 7-19, 2004.
- [Tsoumakas & Katakis, IJDWM07] G. Tsoumakas and I. Katakis. Multi-Label Classification: An Overview. International Journal of Data Warehousing and Mining, 3(3): 1-13, 2007.

References

- [Tsoumakas et al, DMKDH10] G. Tsoumakas, I. Katakis and I. Vlahavas. Mining Multi-label Data. In: Data Mining and Knowledge Discovery Handbook, 2nd edition, Springer, Berlin, 2010.
- [Zhou, JSCT06] Z.-H. Zhou. Multi-Instance Learning from Supervised View. Journal of Computer Science and Technology, 21(5): 800-809, 2006.
- [Zhou & Zhang, NIPS'06] Z.-H. Zhou and M.-L. Zhang. Multi-Instance Multi-Label Learning with Application to Scene Classification. Advances in NIPS, 1609-1616, 2006.
- [Liu, TR06] Y. Liu. Distance Metric Learning: A Comprehensive Survey. Technical Report, Michigan State University, 2006.
- [Shwartz, Thesis07] S. Shalev-Shwartz. Online learning: Theory, Algorithms, and Applications. PhD Thesis, Hebrew university, 2007.
- [Bengio, FTML09] Yoshua Bengio. Learning Deep Architectures for AI. Foundations and Trends in Machine Learning, 2(1): 1-127, 2009.
- [Pan & Yang, TKDE2010] S. Jialin Pan and Q. Yang. A Survey on Transfer Learning. IEEE Transactions on Knowledge and Data Engineering.

References

- [Liu, FTIR09] T.-Y. Liu. Learning to Rank for Information Retrieval. Foundations and Trends in Information Retrieval, 3(3): 225-331, 2009.
- [Evgeniou et al, JMLR05] T. Evgeniou, C. Micchelli and M Pontil. Learning Multiple Tasks with Kernel Methods. Journal of Machine Learning Research, 6:615-637, 2005.
- [Argyriouet et al, NIPS'06] A. Argyriou, T. Evgeniou and M. Pontil. Multi-Task Feature Learning. Advances in NIPS, 41-48, 2007.
- [Kaelbling et al, JAIR96] L. Kaelbling, L. Michael and M. Andrew. Reinforcement Learning: A Survey. Journal of Artificial Intelligence Research, 4:237-285, 1996.
- [Sutton & Barto, Book98] R. Sutton and A. Barto. Reinforcement Learning: An Introduction. MIT Press, Cambridge, MA, 1998.
- [Tong, Thesis01] S. Tong. Active Learning: Theory and Applications. PhD Thesis, Stanford University, 2001.