Artificial Intelligence

School of Electronic and Computer Engineering
Peking University

Wang Wenmin

Contents:

Contents:

Supervised Learning Paradigm

School of Electronic and Computer Engineering
Peking University

Wang Wenmin

Objectives 教学目的

- In this chapter we will discuss in detail about the paradigms that have been proposed in machine learning.
  这一章我们详细讨论针对机器学习所提出的一些范式。

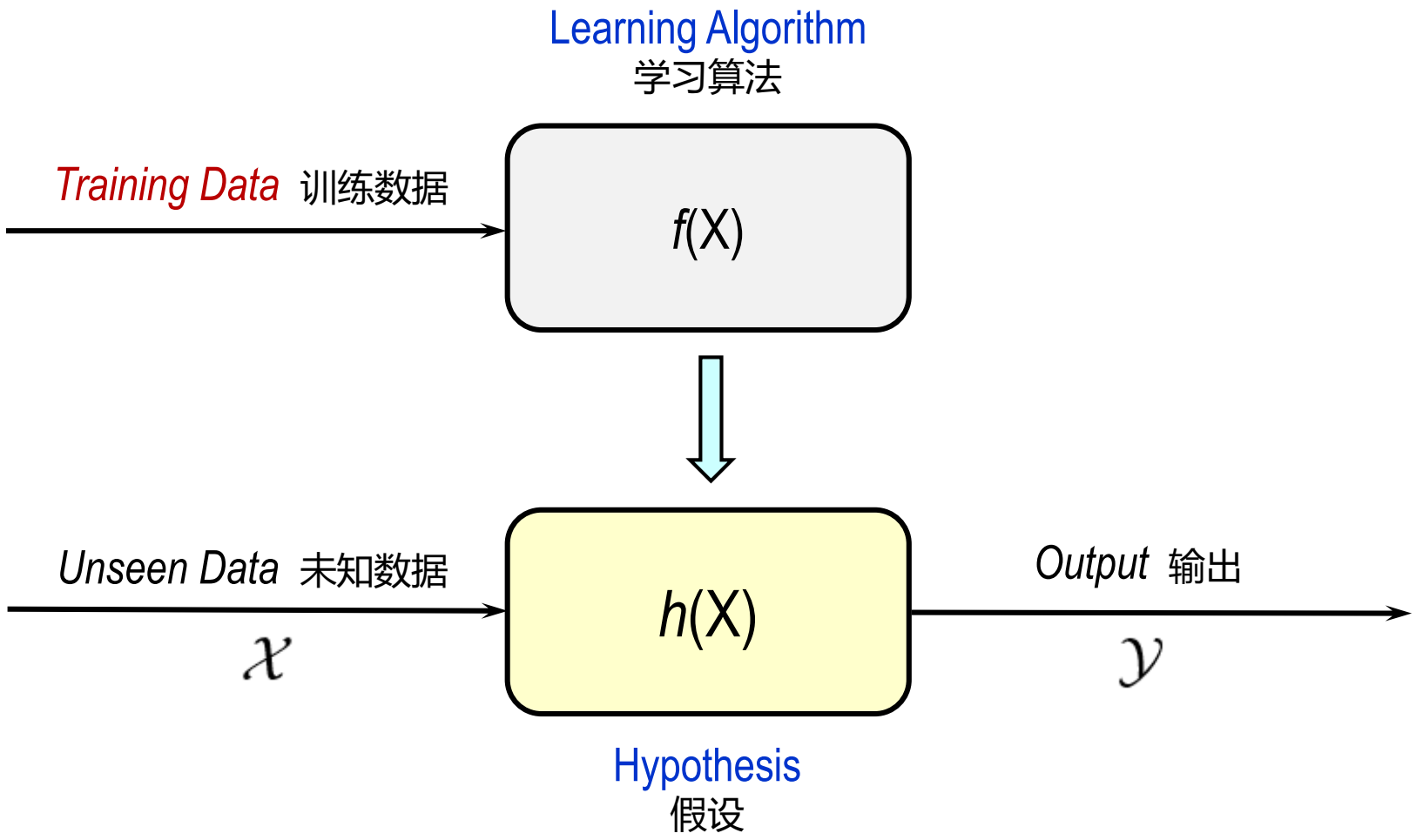
What are Learning Paradigms 什么是学习的范式

- The learning paradigms are used to denote the typical scenarios that are happened in machine learning.
  学习范式用于表示机器学习中发生的典型场景。

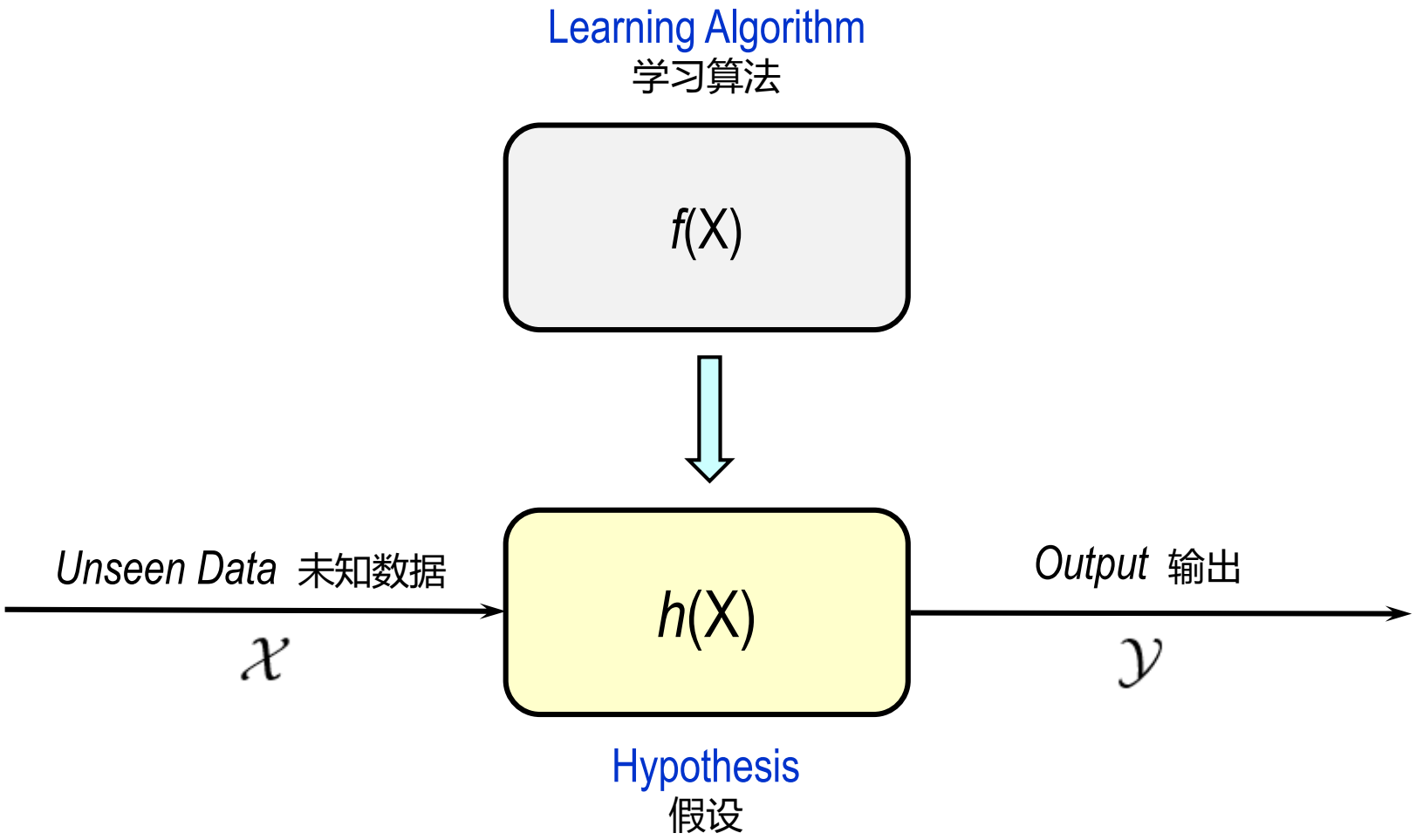
Why Study Learning Paradigms 为什么要研究学习的范式

- Designing an algorithm to solve a learning task may take a different paradigm, such as based on its experience or the interaction with its environment.
  设计一种解决学习任务的算法可能会采用不同的范式，例如基于其经验、或者与其环境的交互。
- Why study the learning paradigms is because it can force you to think about a most appropriate paradigm for the learning task in order to get the best result.
  研读学习范式的意义在于，它可以使你考虑一个最适合该学习任务的范式，以获得最好的结果。

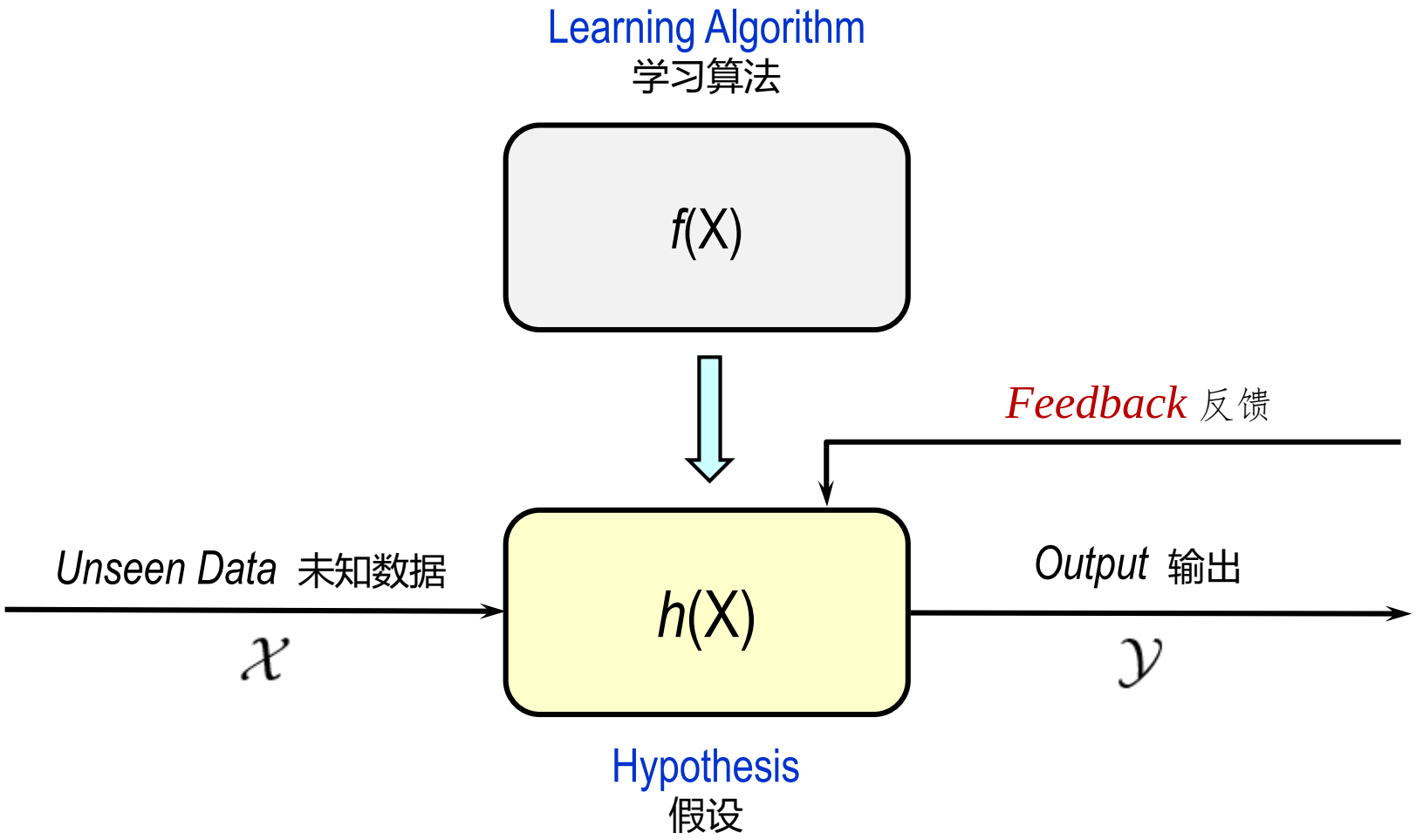
How Does Machine Learning Work 机器学习是如何工作的



How Does Machine Learning Work 机器学习是如何工作的



How Does Machine Learning Work 机器学习是如何工作的



Typical Paradigms in Machine Learning 机器学习中的典型学习范式

Paradigms 范式	Brief Statements 简短描述	Typical Algorithm 典型算法
Supervised 有监督	The algorithm is trained by a set of labeled data, and makes predictions for all unseen points. 算法采用一组标注数据进行训练，再对所有的未知点做出预测。	Support vector machines 支撑向量机
Unsupervised 无监督	The algorithm exclusively receives unlabeled data, and makes predictions for all unseen points. 算法仅接收未标注的数据，再对所有的未知点做出预测。	*k*-means *k*-均值
Reinforcement 强化	The algorithm interacts with environment, and receives an reward for each action. 算法与外部环境交互，每个动作得到一个回报。	Q-learning

11. Paradigms in Machine Learning

Contents:

What is Supervised Learning 什么是有监督学习

- The agent receives a set of labeled examples as training data, and makes predictions for all unseen points.
  智能体接收一组标注的样本作为训练数据，然后对所有的未知点进行推测。
- This approach attempts to generalize a function or mapping from inputs to outputs by training, which can then be used speculatively to generate an output for previously unknown data.
  这种方式试图生成从输入到输出的函数或映射，然后可以将其用于对预先未知的数据生成输出。

It is a way of "teaching" the learning algorithm, like that a "teacher" gives the classes (courses).

这是一种“教”学习算法的方式，就像“老师”讲授课程那样。

What is Supervised Learning 什么是有监督学习

- The training data in supervised learning:
  有监督学习中的训练数据：
  - each training data has a known label as an input data,
    每个训练数据具有一个已知标注作为输入数据，
  - the label is a pair consisting of an input object and a desired output value
    标注是由输入对象和预期输出值组成的对
    (such as spam/not-spam, or a stock price at a time)
    （例如垃圾与非垃圾邮件、或某时刻股票价格）。
- An hypothesis function after training:
  训练后的假设函数：
  - can be used for mapping new unseen data.
    可用于映射新的未知数据。

11.1. Supervised Learning Paradigm

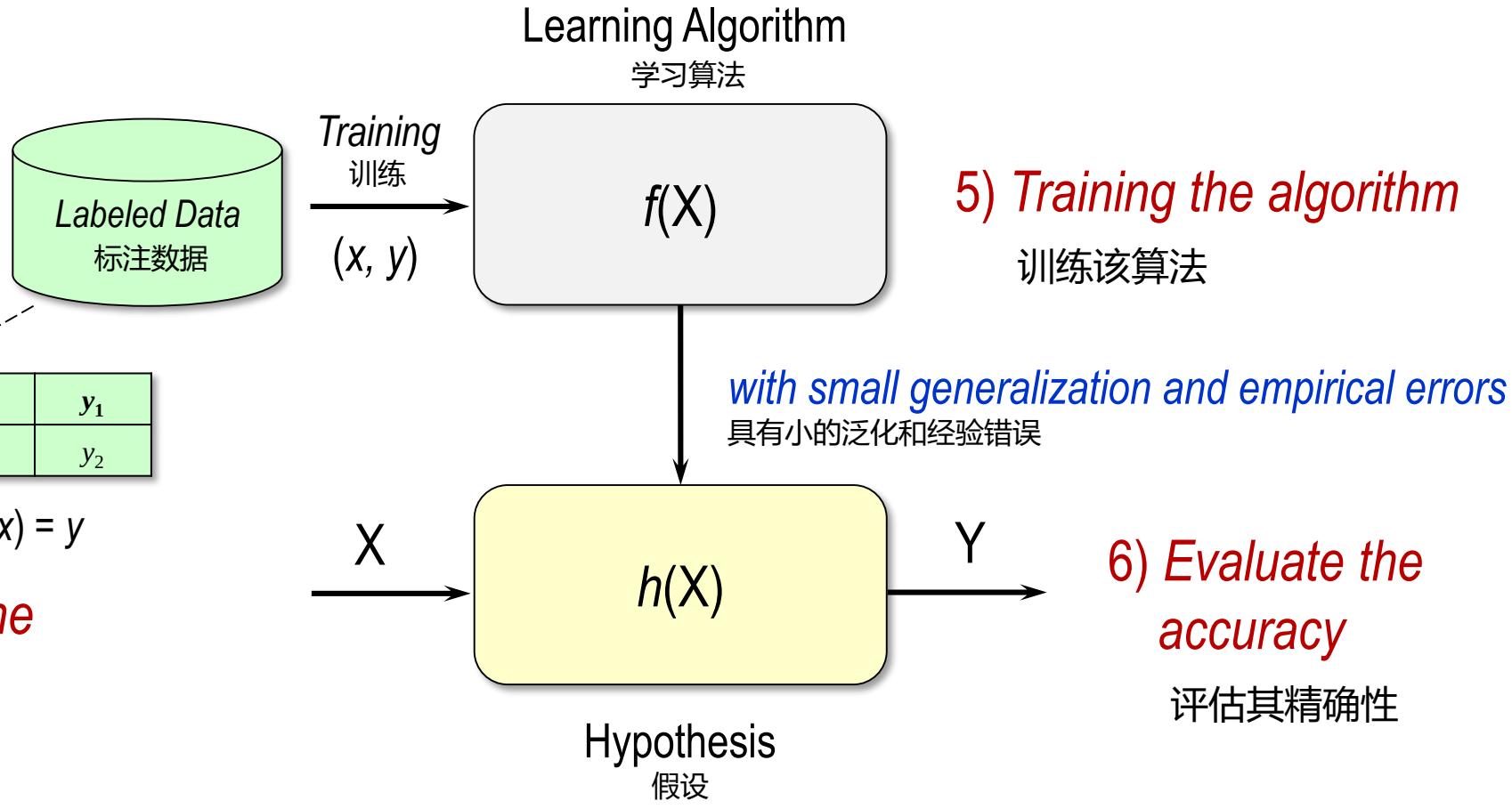
Contents:

Six Steps by Supervised Learning 有监督学习的6个步骤

4) *Determine the algorithm to the task* 设计该任务的算法

3) *Determine the feature extraction approach*

确定特征提取方法



Six Steps by Supervised Learning 有监督学习的6个步骤

1) Determine the training type 确定训练类型

You should decide firstly what kind of data is to be used as a training set.

应该首先确定使用何种数据作为训练集。

- E.g., for handwriting recognition, that may be a single handwritten character, an entire handwritten word, or an entire line of handwriting.

  例如，对于手写体识别，可以是一个手写字符、一个完整的手写单词、或是手写的一行。

2) Gather a training set 收集训练集

The training set needs to be representative of the real-world use of the function.

训练集需要代表实际使用的功能。

- Thus, a set of input objects is gathered and corresponding outputs are also gathered, either from human experts or from measurements.

  因此，由人类专家或者通过测量，筛选出一组输入对象以及对应的输出。

Six Steps by Supervised Learning 有监督学习的6个步骤

3) Determine the feature extraction approach 确定特征提取方法

Typically, there are two kind of approaches to extract the feature from input data:

通常，有两种从输入数据提取特征的方法：

- *handcraft* feature extraction: by some feature descriptor.

  手工特征提取：通过某种特征描述子。

- *automated* feature extraction: by some deep neural network.

  自动特征提取：通过某种深度神经网络。

4) Design the algorithm to the task 设计该任务的算法

This depends on what your task is.

这取决于你的任务是什么。

- E.g., for classification, you may choose to use SVM, decision tree, Softmax, etc.

  例如，对于分类来说，你可以选择使用SVM、决策树、Softmax、等等。

Six Steps by Supervised Learning 有监督学习的6个步骤

5) Training the algorithm 训练该算法

Run the learning algorithm on the gathered training set.

在收集的训练数据集上运行该学习算法。

- Some algorithms require the user to determine certain control parameters.

  某些算法需要用户来确定某些控制参数。

- These parameters may be adjusted by optimizing performance on a subset of the training set.

  这些参数可以通过在训练子集上优化性能来调整。

6) Evaluate the accuracy 评估其精确性

After parameter adjustment and learning, the performance of the resulting function should be measured on a validation set that is separate from the training set.

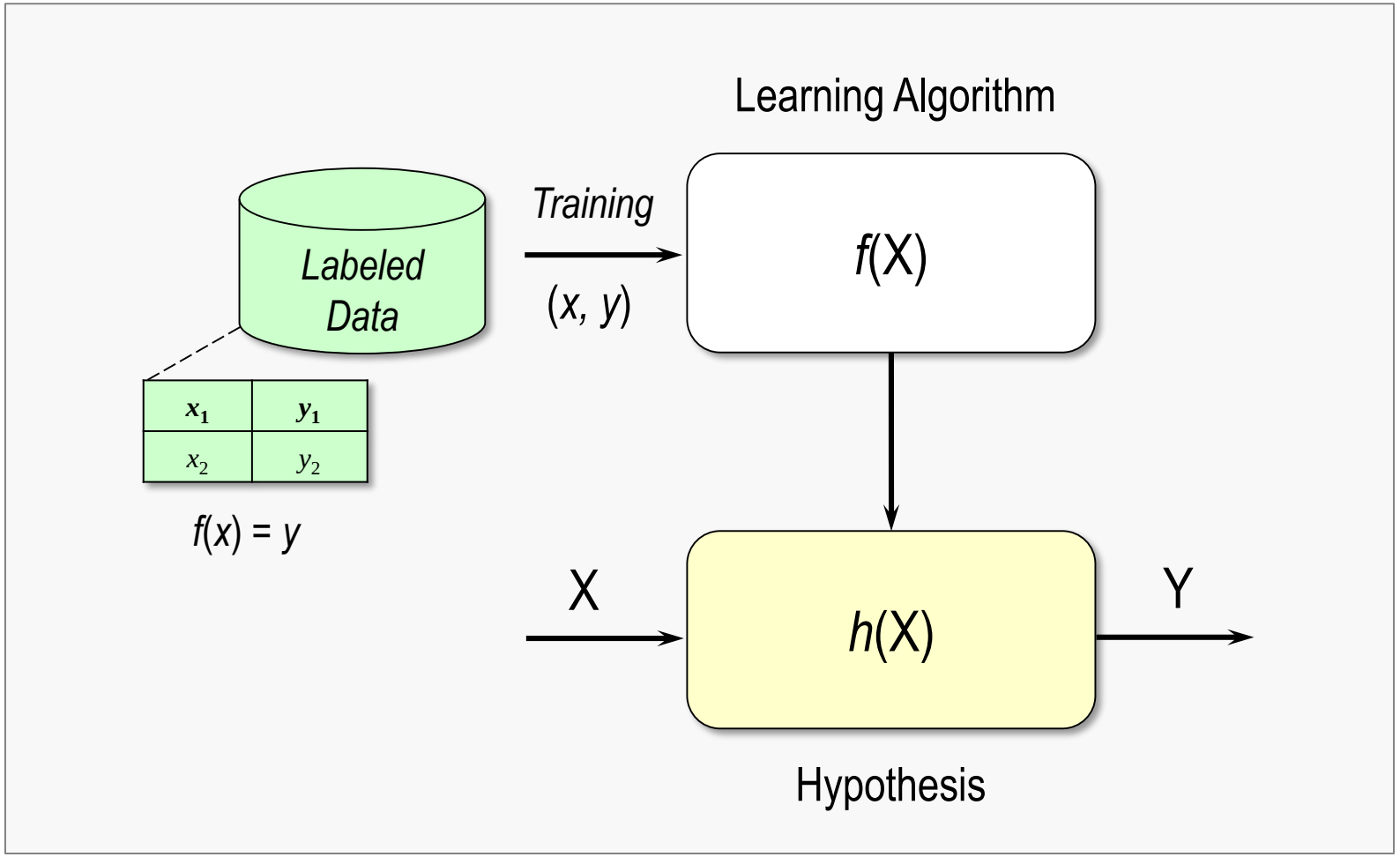
在参数调整和学习之后，应当在（独立于训练集的）验证数据集上对结果函数的性能进行度量。

11.1. Supervised Learning Paradigm

Contents:

Tasks Associated with Supervised Learning 与有监督学习相关的任务

- Classification,
  分类
- Regression,
  回归
- Ranking.
  排名



11.1. Supervised Learning Paradigm

Contents:

A Formal Description for Supervised Learning 一种有监督学习的形式化描述

Let X denote input space, Y denote output space, and *D* an unknown distribution over X × Y.

设X表示输入空间， Y表示输出空间，并且*D*表示X × Y上的一个未知分布。

- Let target labeling function: 设目标标注函数

$$f : \square \rightarrow \square$$

- Training set (a labeled sample set): 训练集（标注的训练样本集）

$$\square = \{(x^{(i)}, y^{(i)}) \mid (x, y) \in \square \times \square, i \in [1, m]\}$$

- Given a hypothesis set *H*, to find a hypothesis $h \in H$ that is the mapping:

  给定假设集*H*，来发现一个假设$h \in H$，满足如下映射：

$$h : \square \rightarrow \square$$

A Formal Description for Supervised Learning 一种有监督学习的形式化描述

- **Classification** 分类

  output space Y is a set of categories.

  输出空间Y是一组类别。

  - **Regression** 回归

    output space Y is a set of real continues numbers.

    输出空间Y是一组连续的实数值。

    - **Ranking** 排名

      output space Y is a set with relative order.

      输出空间Y是一组相对的顺序。

11.1. Supervised Learning Paradigm

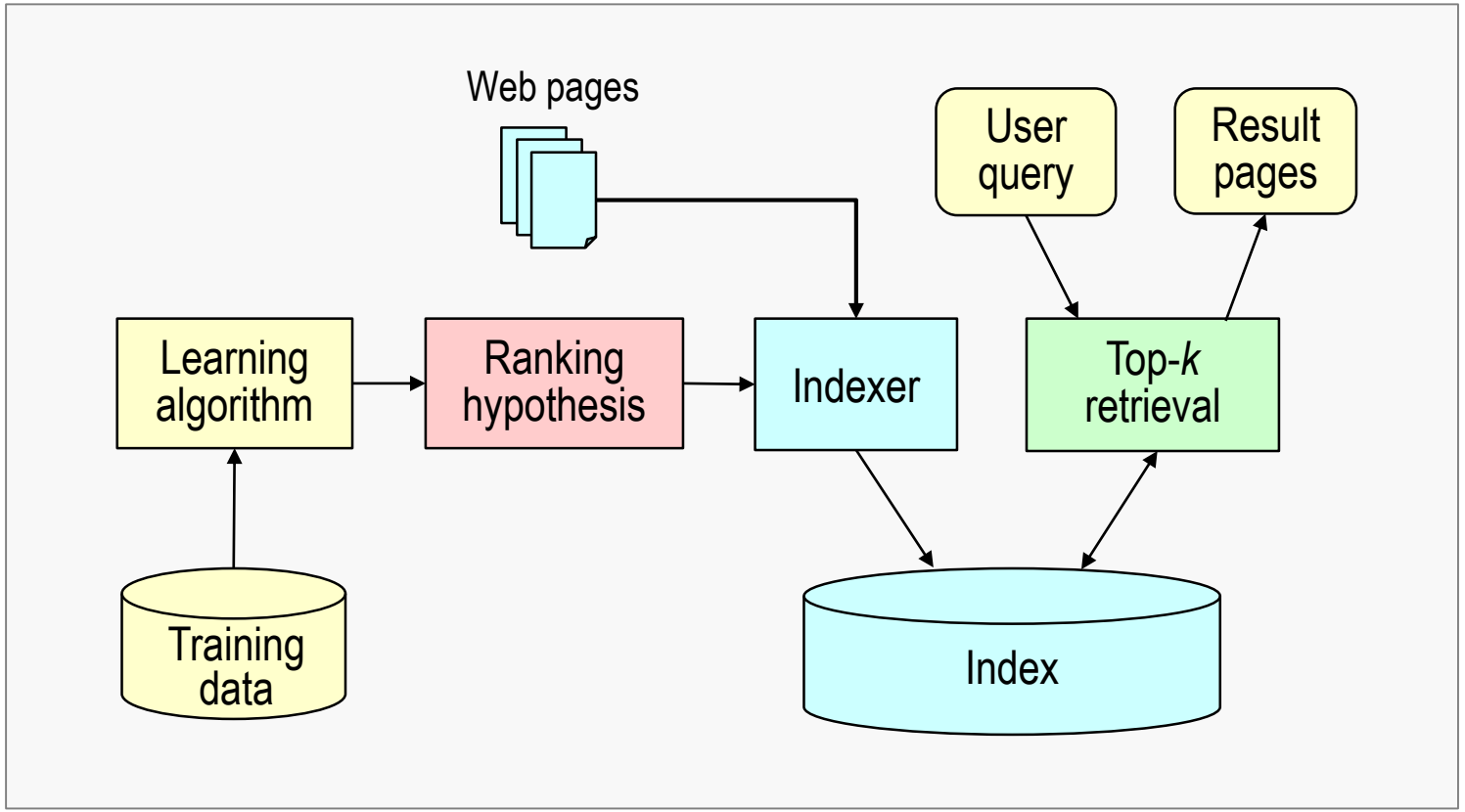
Contents:

Typical Classification and Regression Algorithms 典型的分类与回归算法

Algorithm 算法	Task Types 任务类型	Predictive accuracy 预测精度	Training speed 训练速度
AdaBoost 自适应增强	Either 两者	Higher 高	Slow 慢
Artificial neural network 人工神经网络	Either 两者	Higher 高	Slow 慢
k-Nearest neighbor *k*近邻	Either 两者	Lower 低	Fast 快
Linear regression 线性回归	Regression 回归	Lower 低	Fast 快
Logistic regression 逻辑回归	Classification 分类	Lower 低	Fast 快
Naive Bayes 朴素贝叶斯	Classification 分类	Lower 低	Fast 快
Decision tree 决策树	Either 两者	Lower 低	Fast 快
Random Forests 随机森林	Either 两者	Higher 高	Slow 慢
Support vector machines 支撑向量机	Either 两者	Higher 高	Slow 慢

Typical Ranking Algorithms 典型的排名算法

- AdaRank
- BayesRank
- BoltzRank
- LambdaRank
- RankBoost
- Ranking Refinement
- RankSVM
- PageRank



A concept architecture of a machine-learned search engine.
一种机器学习搜索引擎的概念架构

11.1. Supervised Learning Paradigm

Contents:

Some Applications of Supervised Learning 有监督学习的一些应用

- Object recognition in computer vision 计算机视觉中的物体识别
- Optical character recognition (OCR) 光学字符识别 (OCR)
- Handwriting recognition 手写体识别
- Information retrieval 信息检索
- Learning to rank 学会排名
- Spam detection 垃圾邮件检测
- Speech recognition 语音识别
- Bioinformatics 生物信息学
- Cheminformatics 化学信息学

Some Examples of Supervised Learning 几个有监督学习的例子

Spam Detection 垃圾邮件检测

Mapping email to {Spam, Not Spam}

将电子邮件分为 {Spam, Not Spam}

Digit Recognition 数字识别

Mapping handwriting digit to {0, 1, 2, 3, 4, 5, 6, 7, 8, 9}

将手写体数字映射为 {0, 1, 2, 3, 4, 5, 6, 7, 8, 9}

Price Prediction for Used Cars 二手车价格预测

Mapping a used car to a real price, based on the historical data collected from used car market.

根据二手车市场收集到的历史数据，估算一台二手车的实际价格。

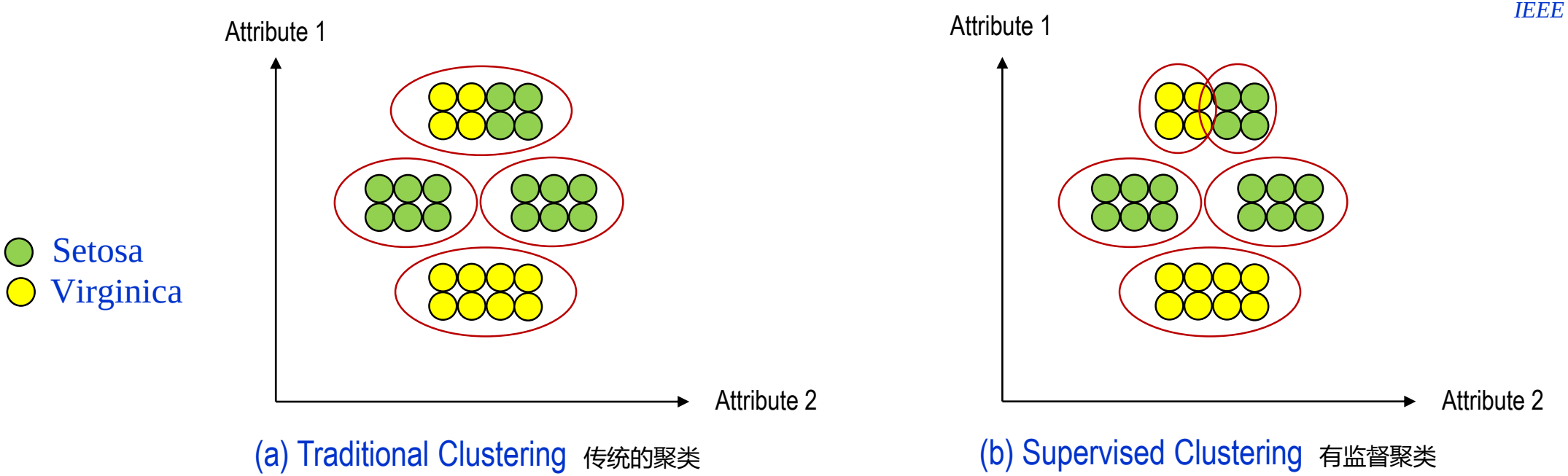
Special Case 特例

Supervised Clustering 有监督聚类

to identify class-uniform clusters that have high probability densities.

识别具有高概率密度的类别统一类聚。

Source: "Supervised clustering - algorithms and benefits", *IEEE ICTAI*, 2004



(a) Traditional Clustering 传统的聚类

(b) Supervised Clustering 有监督聚类

Differences between Traditional Clustering and Supervised Clustering

传统聚类和有监督聚类的差异

11.1. Supervised Learning Paradigm

Contents:

Variants of Supervised Learning 有监督学习的变体

Paradigms 范式	Brief Statements 简介
Semi-supervised learning 半监督学习	A class of supervised learning techniques that also make use of unlabeled data for training. 属于有监督学习算法一类，此外还利用未标记数据进行训练。
Weakly supervised learning 弱监督学习	It aims to learn some information using a limited amount of training examples. 旨在采用有限数量的训练样本来学习一些信息。
One-shot learning 一次性学习	It aims to learn some information from one, or only a few, training examples. 旨在从一个、或仅有的几个训练样本中学习一些信息。
Zero-shot learning 零次性学习	It is able to solve a task despite not having received any training examples of that task. 即使没有得到某个任务的任何训练样本，也能够求解该任务。

11. Paradigms in Machine Learning

Contents:

Thank you for your attention!

AI

Unsupervised Learning Paradigm

School of Electronic and Computer Engineering
Peking University

Wang Wenmin

11.2. Unsupervised Learning Paradigm

Contents:

What is Unsupervised Learning 什么是无监督学习

- The agent exclusively receives unlabeled data, and makes predictions for all unseen points.

  智能体专门接收未标注数据，并对所有的未知点做出预测。

- The objective is discovering commonalities in the data, or reducing the number of random variables under consideration.

  其目标是发现数据中共性的东西，或者减少正在考虑的随机变量的数量。

It is a way of "teaching by itself", without a "teacher".

这是一种“自学”的方式，没有“老师”。

Supervised vs. unsupervised learning 有监督与无监督学习

Supervised learning

有监督学习

- the examples given to the learner are labeled,

  给予学习器的样本是已标注的，
- the examples are used for training the algorithm.

  样本用于训练该算法。

Unsupervised learning

无监督学习

- the examples given to the learner are unlabeled,

  给予学习器的样本是未标注的，
- there is no training process.

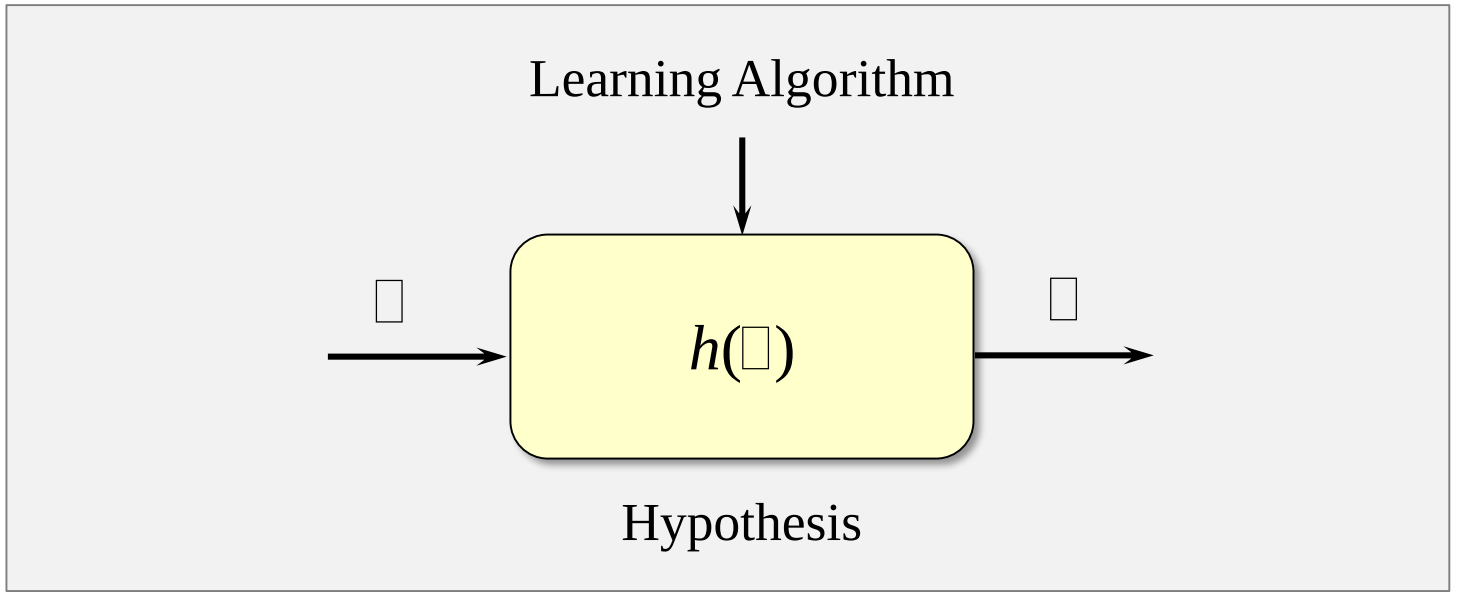
  没有训练过程。

11.2. Unsupervised Learning Paradigm

Contents:

Tasks Associated with Unsupervised Learning 与无监督学习相关的任务

- Clustering 聚类
- Density estimation 密度估计
- Dimensionality reduction 降维



Special Case 特例

- Unsupervised classification 无监督分类

source: http://www.utcmapper.frec.vt.edu

where the outcomes (groupings of pixels with common characteristics) are based on the software analysis of an image without the user providing sample classes.

其输出结果（具有共同特征的像素分组）是基于图像的软件分析，而没有用户提供的样本类。

Classified tree canopy layer in the Virginia Urban Tree Canopy Mapper

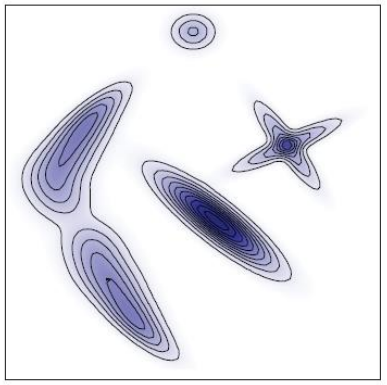
弗吉尼亚城市树冠测绘图中的分类树冠层次

11.2. Unsupervised Learning Paradigm

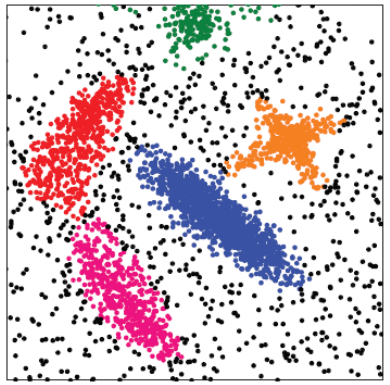
Contents:

Typical Clustering Algorithms 典型的聚类算法

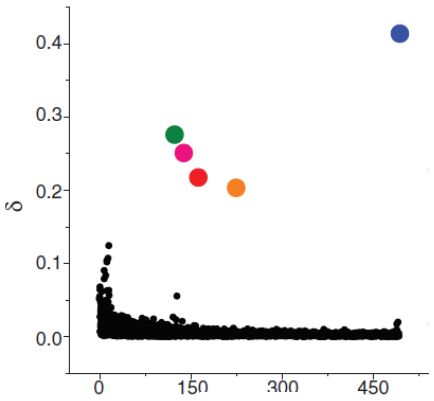
- Single-linkage clustering 单链聚类
- Conceptual clustering 概念聚类
- *k*-means *k*均值
- Fuzzy clustering 模糊聚类
- Clustering by density peaks 密度峰值聚类



(a)



(b)



(c)

Clustering by density peaks
密度峰值聚类

Results for synthetic point distributions.
合成点分布的结果

(a) The probability distribution from which point distributions.
围绕每个簇中心点分布的概率分布图。

(b) Point distributions for samples of 4000 points.
4000点样本的点分布图。

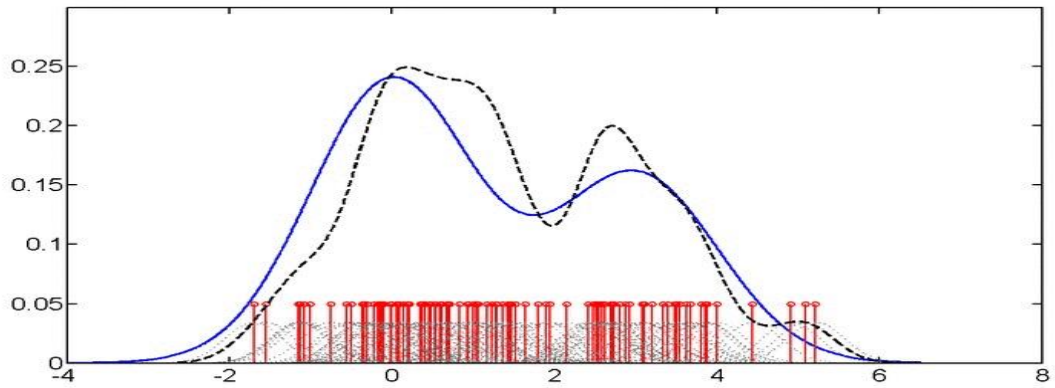
(c) The corresponding decision graph, with the centers colored by cluster.
对应的决策图，彩色点是聚类的中心点。

Source: SCIENCE, Vol. 344, Jun. 27 2014.

Typical Approaches of Density Estimation 典型的密度估计方法

- Kernel density estimation 核密度估计
- Mean integrated squared error (MISE) 平均积分平方误差 (MISE)
- Multivariate kernel density estimation 多变量核密度估计
- Spectral density estimation 谱线密度估计
- Kernel embedding of distributions 分布式核嵌入

Source: https://en.wikipedia.org/wiki/Density_estimation



The density estimation using kernel smoothing

采用核平滑法的密度估计

Typical Dimensionality Reduction Algorithms 典型的降维算法

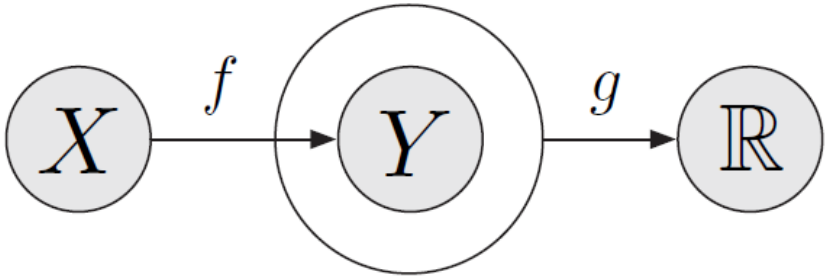
- Principal component analysis (PCA) 主成分分析 (PCA)
- Kernel PCA 核 PCA
- Graph-based kernel PCA 基于图的核 PCA
- Linear discriminant analysis (LDA) 线性判别分析 (LDA)
- Generalized discriminant analysis (GDA) 广义判别分析 (GDA)
- Multi-dimensional Scaling (MDS) 多维尺度分析 (MDS)
- Isometric feature mapping (Isomap) 等距特征映射 (Isomap)
- Locally-linear embedding (LLE) 局部线性嵌入 (LLE)

Case Study: Label-Free Supervision of Neural Networks 无标注有监督神经网络

Motivation 动机

- The neural networks are supervised by specifying *constraints over the output space*, rather than labeled training examples.
  该神经网络是通过指定输出空间的约束来监督的，而不是标注的训练样本。
- The constraints are derived from prior domain knowledge.
  该约束是从先验的领域知识得来的。

Source: "Label-Free Supervision of Neural Networks with Physics and Domain Knowledge", *AAAI* 2017, *best paper*.



Constraint learning aims to recover the transformation *f* without providing labels *y*.
Instead, it looks for a mapping *f* that captures the structure required by *g* (constraint function).
约束学习旨在重获变换函数 *f* 而无需提供标注 *y* 。相反，该方法寻找一个捕捉 *g* (约束函数) 所需结构的映射*f*。

Case Study: Label-Free Supervision of Neural Networks 无标注有监督神经网络

Problem Setup 问题设置

- Traditionally, supervision is to learn a function $f : X \rightarrow Y$. A loss function $l : Y \times Y \rightarrow \mathrm{R}$ is provided, and a mapping is found via:
  传统上，有监督是学习一个函数：$f : X \rightarrow Y$。给定损失函数 $l : Y \times Y \rightarrow \mathrm{R}$，则映射等于：

$$f^* = \arg\min_{f \in \mathcal{F}} \sum_{i=1}^{n} \ell(f(x_i), y_i)$$

- Consider an *unsupervised* approach without labels $f(x) = y$, and optimize for a necessary property of the output, g instead. I.e. search for:
  考虑一个没有标注 $f(x) = y$ 的无监督方法，取而代之对输出 g 的必要特性进行优化。即寻找：

$$\hat{f}^* = \arg\min_{f \in \mathcal{F}} \sum_{i=1}^{n} g(x_i, f(x_i)) + R(f)$$

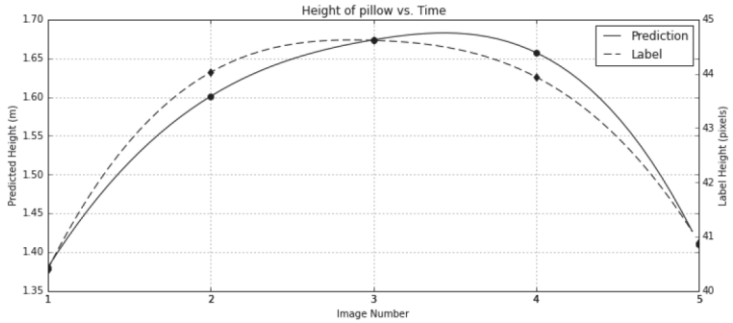
Case Study: Label-Free Supervision of Neural Networks 无标注有监督神经网络

- Experiment: Tracking an object in free fall 实验：跟踪自由落体的对象
  - Plot of object's height over: 绘制对象的高度

$$\mathbf{y}_i = y_0 + v_0(i\Delta t) + a(i\Delta t)^2$$

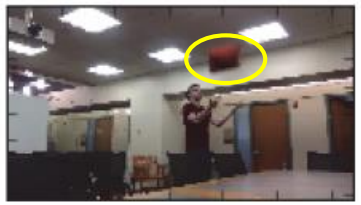
  - Constraint loss: 约束损失

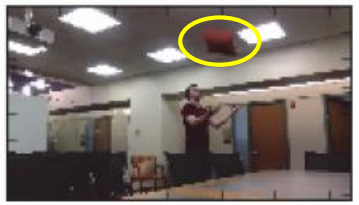
$$g(\mathbf{x}, f(\mathbf{x})) = g(f(\mathbf{x})) = \sum_{i=1}^{N} |\hat{\mathbf{y}}_i - f(\mathbf{x})_i|$$



The parabola by prediction and by label.

通过预测和标注得到的抛物线。





As the pillow is tossed, the height forms a parabola over time.

Can independently predict the pillow's height in each frame without providing labels.

椅垫儿被抛出后，其高度形成了随时间变化的抛物线。可以在没有提供标注的情况下预测每一帧椅垫儿的高度。

11.2. Unsupervised Learning Paradigm

Contents:

Applications of Unsupervised Learning 无监督学习的应用

- Unsupervised Learning can be used to

  无监督学习可用于

  - separate data into groups (clusters),

    将数据区分成若干个组（类聚），

  - find patterns in data to gain valuable information,

    发现数据中的规律从而得到有价值的信息，

  - map high-dimensional data into lower dimensional space.

    将高维数据映射到低维空间。

11.2. Unsupervised Learning Paradigm

Contents:

Yann LeCun's Comment 雅恩·勒昆的点评

- If intelligence was a cake, unsupervised learning would be the cake, supervised learning would be the icing on the cake, and reinforcement learning would be the cherry on the cake.

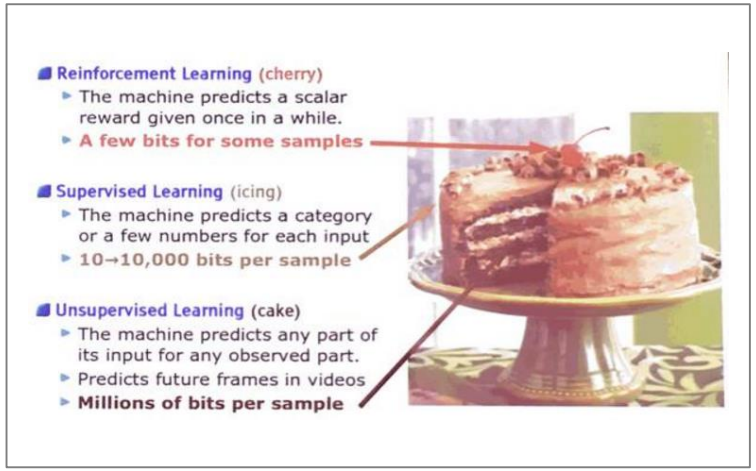
  如果智能是一块蛋糕，无监督学习就是这块蛋糕，有监督学习则是蛋糕上的糖霜，而强化学习是蛋糕上的樱桃。

- We know how to make the icing and the cherry, but we don't know how to make the cake.

  我们知道如何制造糖霜和樱桃，但不知道如何去做这块蛋糕。

- We need to solve the unsupervised learning problem before we can even think of getting to true AI.

  我们需要解决无监督学习问题，然后我们才可以考虑去得到真正的AI。



Source: Yann LeCun, "*Predictive Learning*", invited talk, NIPS 2016

Thank you for your attention!

AI

Reinforcement Learning Paradigm

School of Electronic and Computer Engineering
Peking University

Wang Wenmin

11. Paradigms in Machine Learning

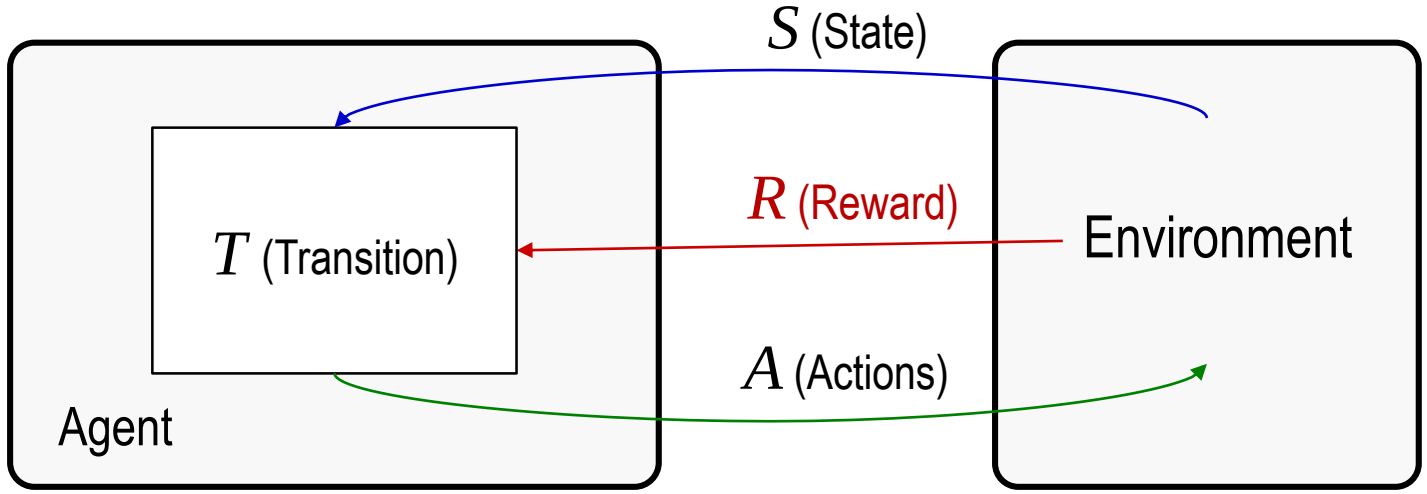
Contents:

11.3. Reinforcement Learning Paradigm

Contents:

What is Reinforcement Learning 什么是强化学习

- In reinforcement learning (RL), the learner is a decision-making agent, that takes actions in an environment and receives rewards for its actions.
  在强化学习中，其学习器是一个决策制定智能体，在环境下采取行动并获得这些动作的回报。
- After a set of trial-and-error runs, the agent should learn the best policy.
  经过一系列试错运行之后，该智能体能够学到最优策略。
- The policy is to maximize his reward over a course of actions and iterations with the environment.
  该策略是经过一个阶段的动作以及与环境的交互之后，使其回报最大化。



What is Reinforcement Learning 什么是强化学习

- Reinforcement Learning is inspired by behaviorist psychology.
  强化学习的灵感来自于行为心理学。
- Concerned with how agents take actions in an environment so as to maximize some notion of cumulative reward.
  关注于智能体如何在环境中采取行动，为了使累积回报最大化。
- Due to its generality, the problem is studied in many other disciplines, such as:
  由于其普遍性，许多其他学科都研究这一问题，例如：
  - game theory, control theory, operations research, information theory,
    博弈论、控制论、运筹学、信息论、
  - simulation-based optimization, multi-agent systems, swarm intelligence,
    仿真优化、多智能体系统、群体智能、
  - statistics and genetic algorithms.
    统计学和遗传算法。

Formalization of Reinforcement Learning 强化学习的形式化

- Reinforcement learning consists of: 强化学习包含
  - a set of agent states, 一组智能体的状态，$s_t \in S$;
  - a set actions of the agent, 一组智能体的动作，$a_t \in A$;
  - a transition from states to actions, 一个从状态到动作的转换函数，$T(s_t, a_t, s_{t+1})$;
  - a reward function, 一个回报函数，$R(s_t, a_t, s_{t+1})$.
- To look for a policy, 寻找一个策略，$\pi(s_t)$.
- Don't know *T* or *R* 尚未知道*T*或*R*
  - I.e. don't know which states are good or what the actions do.
    即，不知道哪个状态好或者要做什么动作。
  - Must actually try actions and states out to learn.
    必须实际去尝试要学习的行动和状态。

Supervised vs. Unsupervised vs. Reinforcement Learning 三种范式之比较

Supervised learning
有监督学习

- Input/output pairs are presented by labeled data (training examples).
  通过标注数据（训练样本）提供输入和输出对儿。
- *Learn-by-examples*
  从样本中学习

Unsupervised learning
无监督学习

- To find the structure hidden in collections of unlabeled data.
  发现无标注数据集中隐藏的结构。
- *Learning-by-itself*
  自我学习

Reinforcement learning
强化学习

- Input/output pairs are never presented, focus on online performance.
  不提供输入和输出对儿，专注于在线的性能优化。
- *Online-learning*
  在线学习

11.3. Reinforcement Learning Paradigm

Contents:

Types of Reinforcement Learning 强化学习的类型

1) Model-based 基于模型

building a model of the environment. 构建环境的模型。

- First acting in Markov decision process (MDP) and learning *T*, *R*;
  首先以马可夫决策过程方式动作，并学习*T*和*R*；
- Then doing value iteration or policy iteration with learned *T*, *R*.
  然后用学习的*T*和*R*进行数值迭代或策略迭代。

2) Model-free 无模型

learning a policy without any model. 学习策略而没有任何模型。

- Bypassing the need to learn *T*, *R*, using direct evaluation policy.
  避开学习*T*和*R*的过程，采用直接评估策略。
- Prediction-based temporal difference (TD) methods.
  基于预测的时间差分 (TD) 法。

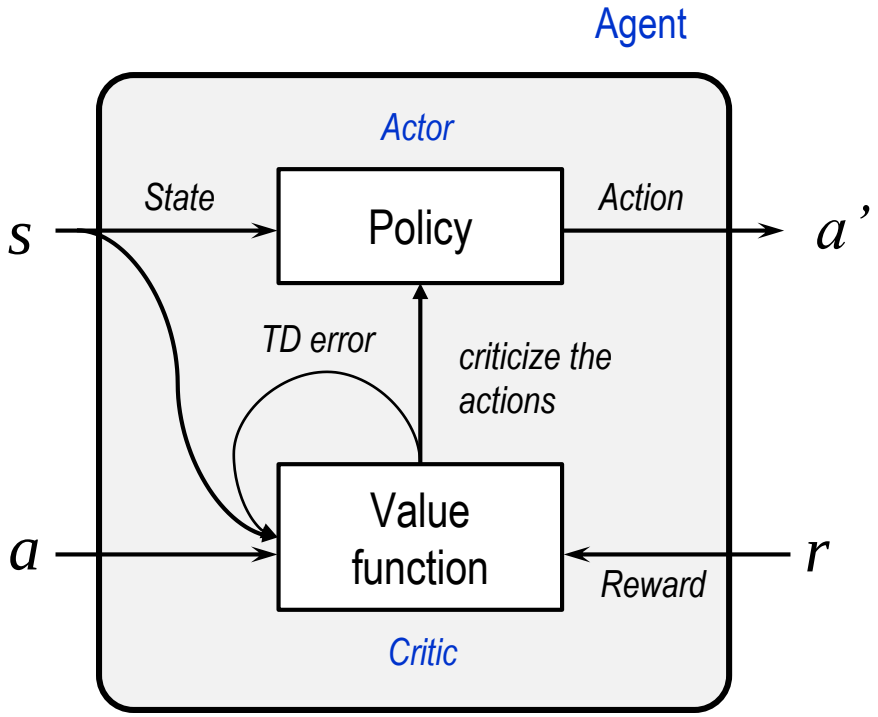
1) Model-based Reinforcement Learning 基于模型的强化学习

- Idea 思想
  - Learning the model empirically through experience. And solving for values as if the learned model were correct.
    通过实践经验学习模型。若学到的模型正确，则用于数值求解。
- Simple empirical model learning 简单的经验模型学习
  - Counting outcomes for each s, a.
    对每个s和a，对结果进行计数。
  - Normalizing to give estimate of $T(s_t, a_t, s_{t+1})$.
    对给定的估计$T(s_t, a_t, s_{t+1})$ 做正则化处理。
  - Discovering $R(s_t, a_t, s_{t+1})$ when we experience (s_t, a_t, s_{t+1}).
    当实践 (s_t, a_t, s_{t+1}) 时，去发现$R(s_t, a_t, s_{t+1})$。
- Solving Markov decision process with the learned model.
  用学到的模型求解马可夫决策过程。

2) Model-free Reinforcement Learning 无模型强化学习

Actor-Critic methods 动作者·评判者方法

- The TD version of Policy Iteration (On-policy).
  策略迭代 (On-policy) 的时间差分版。
- A structure to explicitly represent policy independent of value function.
  一种明确表示独立于价值函数的策略的结构。
- Policy (actor), is used to select actions.
  策略 (动作者) 用于选择动作。
- Value function (critic), used to evaluate actions made by actor.
  价值函数 (评判者) 用于评估动作者所完成的动作。



Actor-critic methods
动作者·评判者方法

TD error: $\delta_t = r_{t+1} + \gamma V(S_{t+1}) - V(S_t)$

Preference: $p(s_t, a_t) \leftarrow p(s_t, a_t) + \beta\delta_t$

2) Model-free Reinforcement Learning 无模型强化学习

Q-learning

- The TD version of Value Iteration (Off-policy).
  价值迭代 (Off-policy) 的时间差分版。
- Incrementally estimate Q-values for actions, based on rewards and Q-value function.
  基于回报值和Q-value函数，递增估计动作的Q值。
- Update rule is a variation of TD learning, using Q-values and a built-in max-operator over the Q-values of the next state:
  更新规则是一种时间差分学习的变体，采用Q值与内置的下个状态Q值的最大运算符：

$$Q(s_t, a_t) = \sum_a T(s_t, a_t, s_{t+1}) \left[R(s_t, a_t, s_{t+1}) + \gamma \max_{a_{t+1}} Q(s_{t+1}, a_{t+1})\right]$$

- Sample-based, action-value function Q will be learned.
  学习到基于样本的、动作-值函数Q。

11.3. Reinforcement Learning Paradigm

Contents:

New Algorithms of Reinforcement Learning 强化学习的新算法

Deep Q-Network (DQN)

深度Q-Network

- CNN + Q-Learning (NIPS'13, Nature'15).

  将CNN与Q-Learning相结合。

Deterministic Policy Gradients (DPG)

确定性策略梯度

- Estimate much more efficiently than usual stochastic policy gradient (ICML'14).

  与常用的随机策略梯度相比，可以更有效地进行估计。

Asynchronous Advantage Actor-Critic (A3C)

异步优势动作者·评判者

- A variant of actor-critic method, using asynchronous gradient descent for optimization of DNN controllers. (arXiv:1602.01783)

  一种作者·评判者的变体，采用异步梯度下降来优化DNN控制器。

New Algorithms of Reinforcement Learning 强化学习的新算法

UNsupervised REinforcement and Auxiliary Learning (UNREAL)

无监督强化及辅助学习

- For the environments containing a much wider variety of possible training signals (arXiv:1611.05397).

  针对包含更广泛的各种可能的训练信号环境。

- It also maximize many other pseudo-reward functions simultaneously.

  还可以同时将许多其它的伪回报函数进行最大化。

Neural Episodic Control (NEC)

神经情景控制

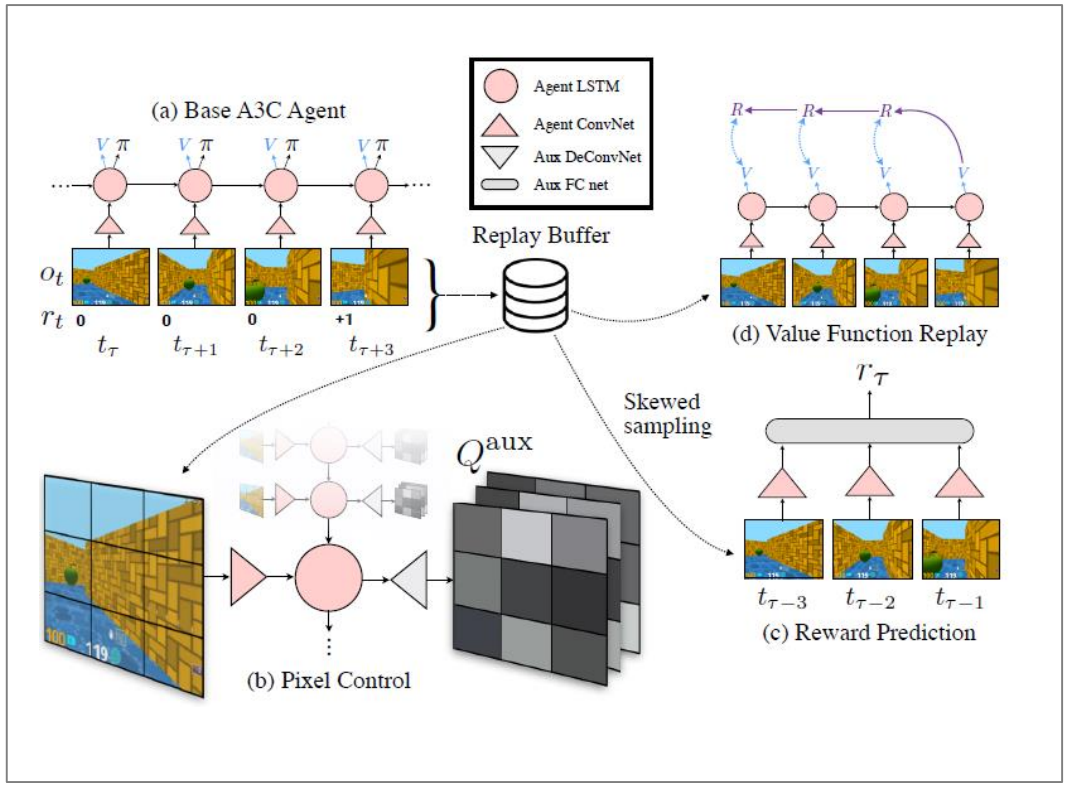
- Can rapidly assimilate new experiences and act upon them (arXiv:1703.01988).

  可以迅速地吸收新的经验，并且对其采取行动。

Case Study: UNREAL (Unsupervised Reinforcement and Auxiliary Learning)

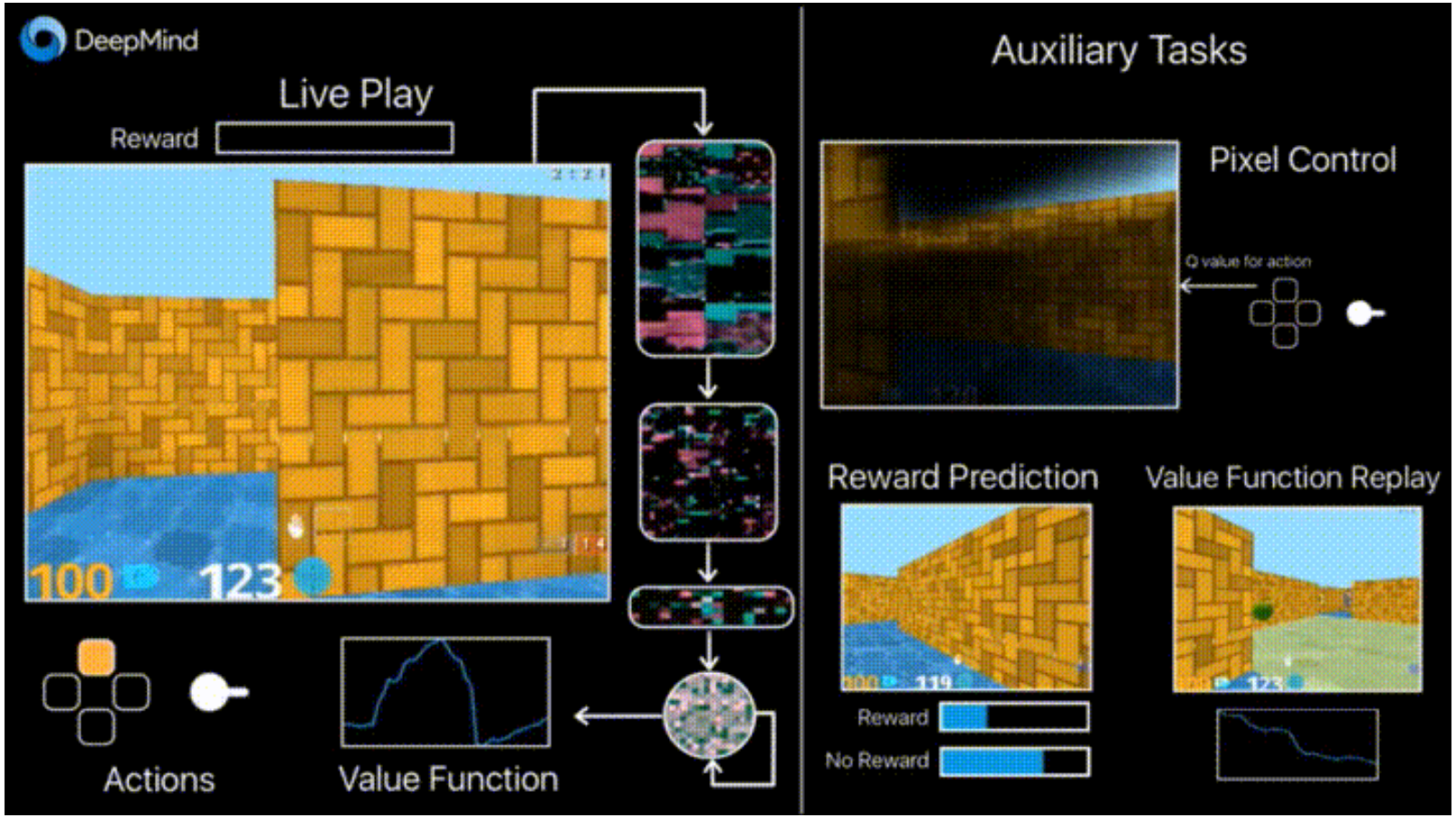
- (a) Base A3C Agent 基础A3C智能体
  a CNN-LSTM agent trained *on-policy* with A3C loss.
  一个CNN-LSTM智能体，经过A3C损失on-policy训练。
- (b) Pixel Control 像素控制
  training auxiliary policies Q^{aux} to maximise change in pixel intensity of different regions.
  训练辅助策略Q^{aux}，使不同区域像素强度变化达到最大化。
- (c) Reward Prediction 回报预测
  given three recent frames, predict the reward that will be obtained in next unobserved timestep.
  给定三个最近的帧，预测将在下一个未观测时阶获得的回报。
- (d) Value Function Replay 价值函数回放
  further training of value function using agent network to promote faster value iteration.
  进一步训练价值函数，采用智能体网络来推进迅速价值迭代。

Source: "Reinforcement learning with unsupervised auxiliary tasks", arXiv:1611.05397, DeepMind



Overview of the UNREAL agent.
UNREAL智能体概览

Case Study: UNREAL (Unsupervised Reinforcement and Auxiliary Learning)



3D Labyrinth on Atari, averaging 880% expert human performance.

Atari上的3D迷宫游戏，平均性能达到人类玩家的880%

11.3. Reinforcement Learning Paradigm

Contents:

Typical Applications of Reinforcement Learning 强化学习的典型应用

- Robots 机器人
  - Robotic arms 机器人手臂

    be controlled to find the most efficient motor combination.

    控制得到最有效的电机组合。
  - Robot navigation 机器人导航

    collision avoidance behavior can be learned by negative feedback.

    可通过负反馈来学会碰撞躲避行为。
- Computer games 计算机游戏
  - Backgammom，西洋双陆棋
  - Chess，国际象棋
  - Go．围棋

Thank you for your attention!

AI

Relations and Other Paradigms

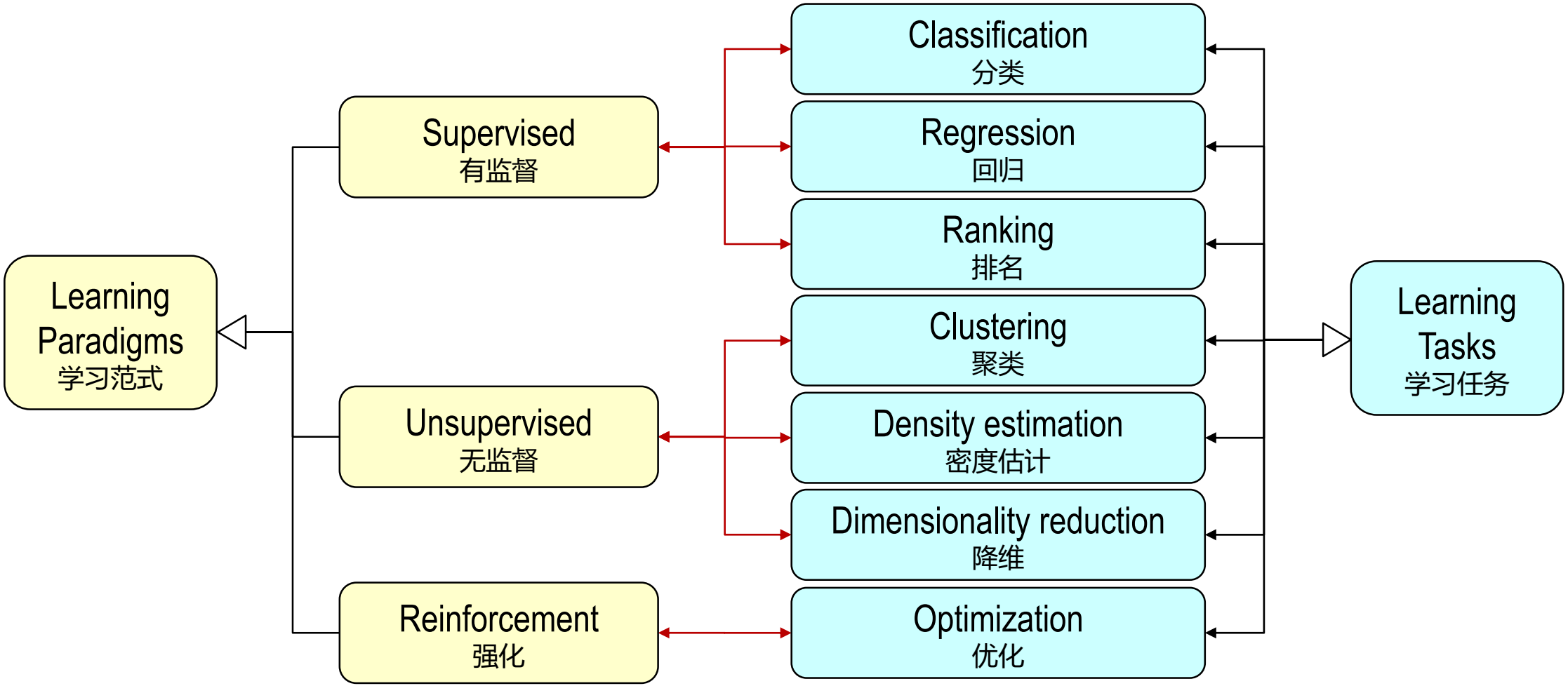
School of Electronic and Computer Engineering
Peking University

Wang Wenmin

11. Paradigms in Machine Learning

Contents:

Relations between Learning Tasks and Paradigms 学习任务与范式的关系



Other Paradigms in Machine Learning 机器学习中的其它范式

Paradigms 范式	Brief Statements 简介
Ensemble learning 集成学习	Combining many weak learners to produce a strong learner. 将多个弱学习器组成一个强学习器。
Learning to learn 学会学习	Learning the inductive bias based on previous experience. 基于先前的经验学习归纳偏差。
Transfer learning 迁移学习	Applying storing knowledge to a different but related problem. 将已有的知识用于不同但相关的问题。
Adversarial learning 对抗式学习	In an adversarial manner (zero sum game) to generate data mimicking some distribution. 以一种对抗性方式（即零和博弈）来生成模仿某种分布的数据。
Collaborative learning 协同式学习	In some collaborative manner (e.g., non-zero sum, win-win) to obtain the desired outputs. 以某种协同式（如非零和博弈、双赢）来得到所期待的输出结果。

Thank you for your attention!

AI