Applications and Terminologies

School of Electronic and Computer Engineering
Peking University

Wang Wenmin

9. Perspectives about Machine Learning

Contents:

- 9.1. What is Machine Learning
- 9.2. History of Machine Learning
- 9.3. Why Different Perspectives
- 9.4. Three Perspectives on Machine Learning
- 9.5. Applications and Terminologies

Application Fields of Machine Learning 机器学习的应用领域

- Machine Perception 机器感知
- Computer Vision 计算机视觉
- Video Analysis 视频分析
- Pattern Recognition 模式识别
  - Face/Speech/Fingerprint Recognition 人脸/语音/指纹识别
  - Optical Character Recognition (OCR) 光学字符识别 (OCR)
  - Handwriting Recognition 手写体识别
- Game Playing 玩游戏
- Natural Language Processing 自然语言处理
- Information Retrieval 信息检索

Application Fields of Machine Learning 机器学习的应用领域

- Text or Document Classification 文本与文档分类
  - (e.g. Spam Email Detection) (例如垃圾邮件检测)
- Recommender Systems 推荐系统
- Ad Placement 广告配置
- Credit Scoring 信用评分
- Fraud Detection 欺诈检测
- Stock Trading 股票交易
- Drug Design 新药设计
- Medical Diagnosis 医学诊断
- Robotics 机器人学

Some Terminologies in Machine Learning 机器学习中的一些术语

☐ Samples 样本

- Items or instances of data used for learning or evaluation.
  用于学习或评估的数据项或实例。

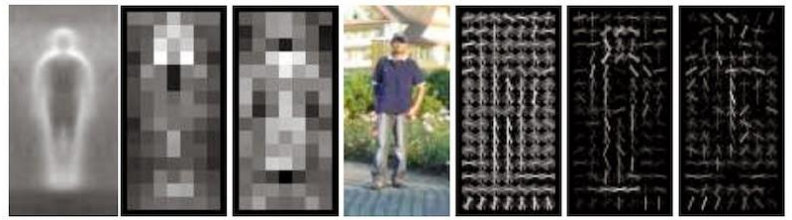
☐ Features 特征

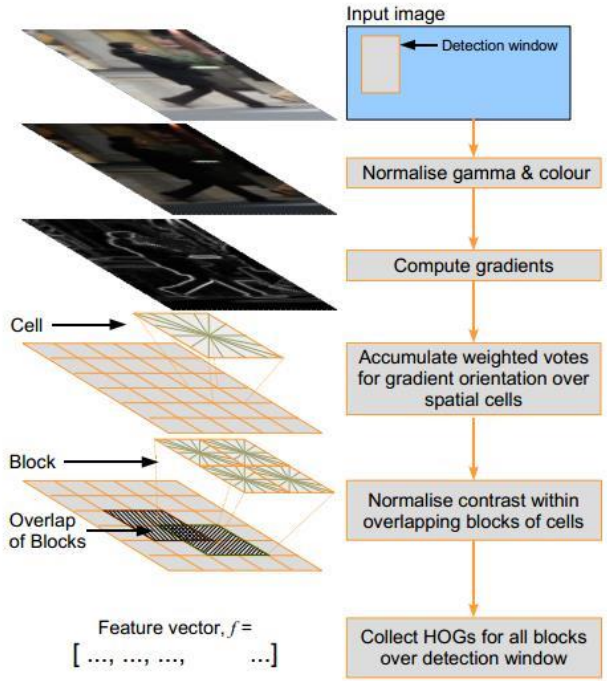
- The set of attributes, often represented as a vector associated to a sample:
  属性集，通常表示为与样本相关的向量：
  - Handcrafted features: 手工式特征
    e.g., SIFT, HOG, SURF, LBP, GLOH, LESH, CENTRIST.
    例如，SIFT、HOG、SURF、LBP、GLOH、LESH、CENTRIST。
  - Learned features: 学习式特征
    e.g., by convolutional neural network.
    例如，通过卷积神经网络。

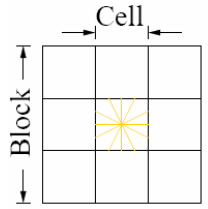
Some Terminologies in Machine Learning 机器学习中的一些术语

☐ Handcrafted Features 手工式特征

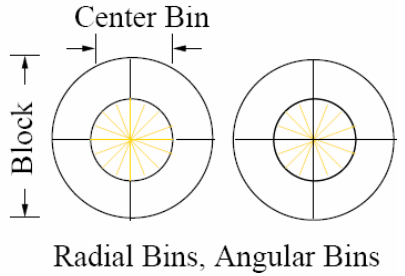
- HOG (Histogram of Oriented Gradients)
  HOG (定向梯度直方图)
- Similar to SIFT(Scale-Invariant Feature Transform), but improved accuracy.
  与SIFT (尺度不变特征变换) 类似，但改善了精度。
- By distribution of intensity gradients or edge directions.
  按照强度梯度或边缘方向分布。
- 64×128 detection window.
  64×128个检测窗口。







R-HOG

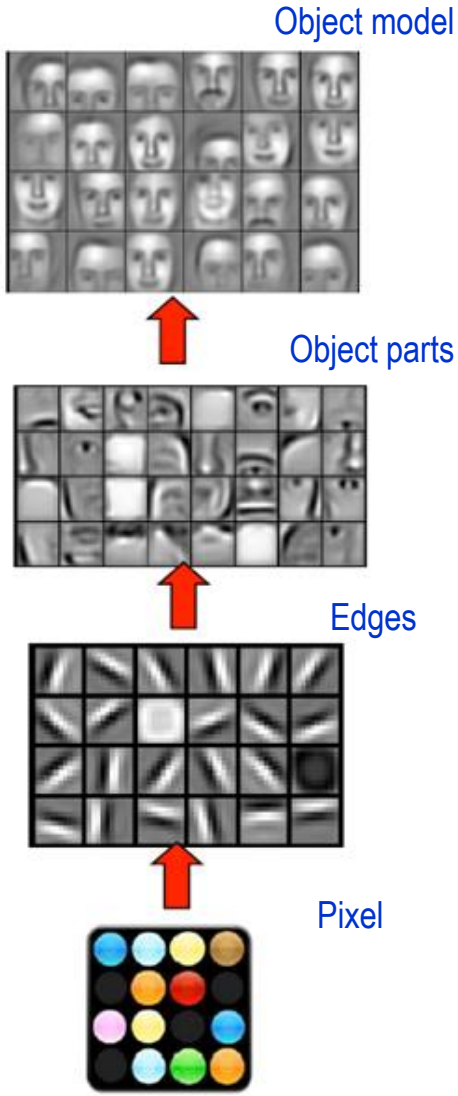


C-HOG

Some Terminologies in Machine Learning 机器学习中的一些术语

☐ Learned Features 学习式特征

- Humans can learn to see efficiently. Because brains are deep, with many layers of processing.
  人类可以有效地学会观察。因为大脑是深度的，具有许多处理层次。
- Some algorithms for such deep architectures, can produce features from raw data for visual recognition.
  具有这种深度架构的算法，能从原始数据中生成视觉认知的特征。
- Feature learning also be called representation learning.
  特征学习也被称为表示学习。
- Understanding deep learning will enable us to build more intelligent machines for visual recognition.
  理解深度学习将使我们能够构建更智能的视觉认知机器。



Some Terminologies in Machine Learning 机器学习中的一些术语

☐ Labels 标记

- Values or categories assigned to samples.
  在样本上指定的值或类别。
- In classification problems, samples are assigned specific categories.
  分类问题中，样本被指定特定的类别。
- In regression problems, items are assigned real-valued labels.
  回归问题中，项被指定为实值的标记。

☐ Training sample 训练样本

- Samples used for training learning algorithm.
  用于训练学习算法的样本。
- In spam problem, the training sample consist of a set of email samples along with their associated labels.
  对于垃圾邮件问题，训练样本由一组邮件样本以及相关标签组成。

Some Terminologies in Machine Learning 机器学习中的一些术语

Validation sample 验证样本

- Samples used to tune the parameters of a learning algorithm when working with labeled data.
  用于在使用标记数据时调整学习算法参数的样本。
- Learning algorithms typically have one or more free parameters, and validation sample is used to select appropriate values for these model parameters.
  学习算法通常具有一个或多个自由参数，因而验证样本用于为这些模型参数选择适当的值。

Test sample 测试样本

- Samples used to evaluate the performance of a learning algorithm.
  用于评估学习算法性能的样本。
- These predictions are then compared with the labels of the test sample to measure the performance of the algorithm.
  然后将这些预测与测试样本的标签进行比较，以衡量算法的性能。

Some Terminologies in Machine Learning 机器学习中的一些术语

☐ Loss function 损失函数

- To measure the difference, or loss, between a predicted label and a true label.
  用于度量预测标签和真实标签之间差异或损失。
- Denoting the set of all labels as □ and the set of possible predictions as □', a loss function L is a mapping:
  将所有的标签集表示为□、并且可能的预测集为□'，则损失函数L为映射：

$$L: \square \times \square' \rightarrow \mathbb{R}_+$$

☐ Hypothesis set 假设集

- A set of functions mapping features to the set of labels □.
  将特征映射为标签□的函数集。
- For example, the following are a set of functions mapping email features to:
  例如，映射电子邮件特征的函数集如下：

$$\square = \{\text{spam}, \text{non}-\text{spam}\}.$$

Some Terminologies in Machine Learning 机器学习中的一些术语

☐ Abstraction 抽象

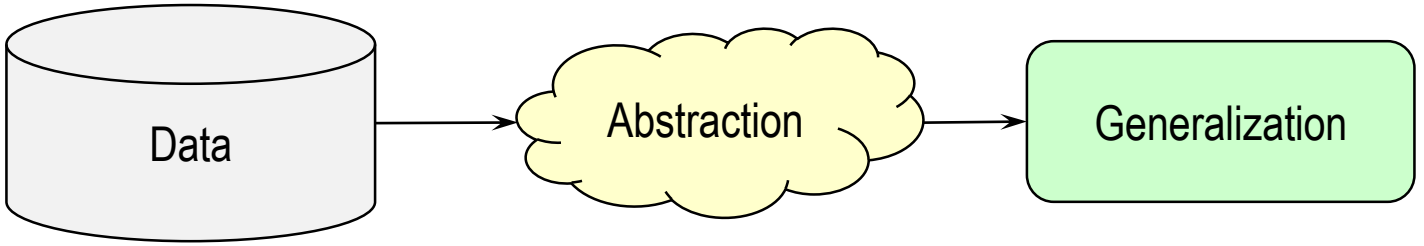
- It involves the translation of data into broader representations.

  其含义是将数据转化为更广泛的表示。

☐ Generalization 泛化

- It describes the process of turning abstracted knowledge into a form that can be utilized for action. It is also the ability of a learning algorithm to perform accurately on unseen samples after having experienced a learning data set.

  它形容将抽象知识转化为可用于动作形式的过程。它也是学习算法具有学习数据集的经验后，可以对未知样本正确地进行处理的能力。



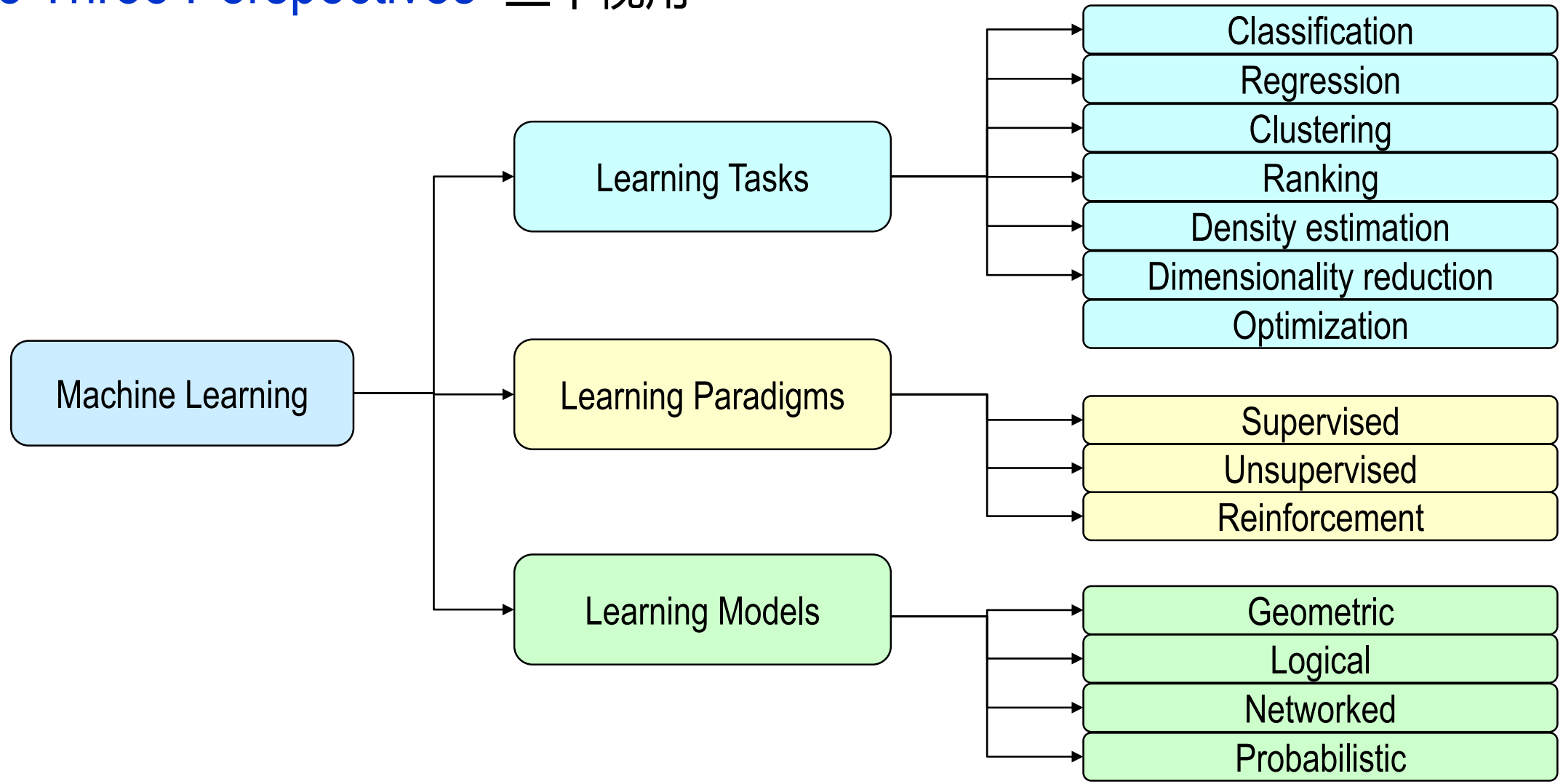
Some Notations in This Course 本课程的一些符号

Notation 符号	Description 说明	
$\mathbb{R}$	Set of real numbers	实数集
$\mathbb{R}_+$	Set of non-negative real numbers	非负实数集
$\mathbb{R}^n$	Set of n-dimensional real-valued vectors	n维实值向量集
$[a, b]$	Closed interval between a and b	a和b之间的闭区间
(a, b)	Open interval between a and b	a和b之间的开区间
$\mathbb{N}$	Set of natural numbers, i.e., $\{0, 1, \ldots\}$	自然数集，即：$\{0, 1, \ldots\}$
$\mathcal{X}$	An arbitrary set	任意集合
$\mathcal{X}$	Input space	输入空间
$\mathcal{Y}$	Target space	目标空间
H	hypothesis set	假设集

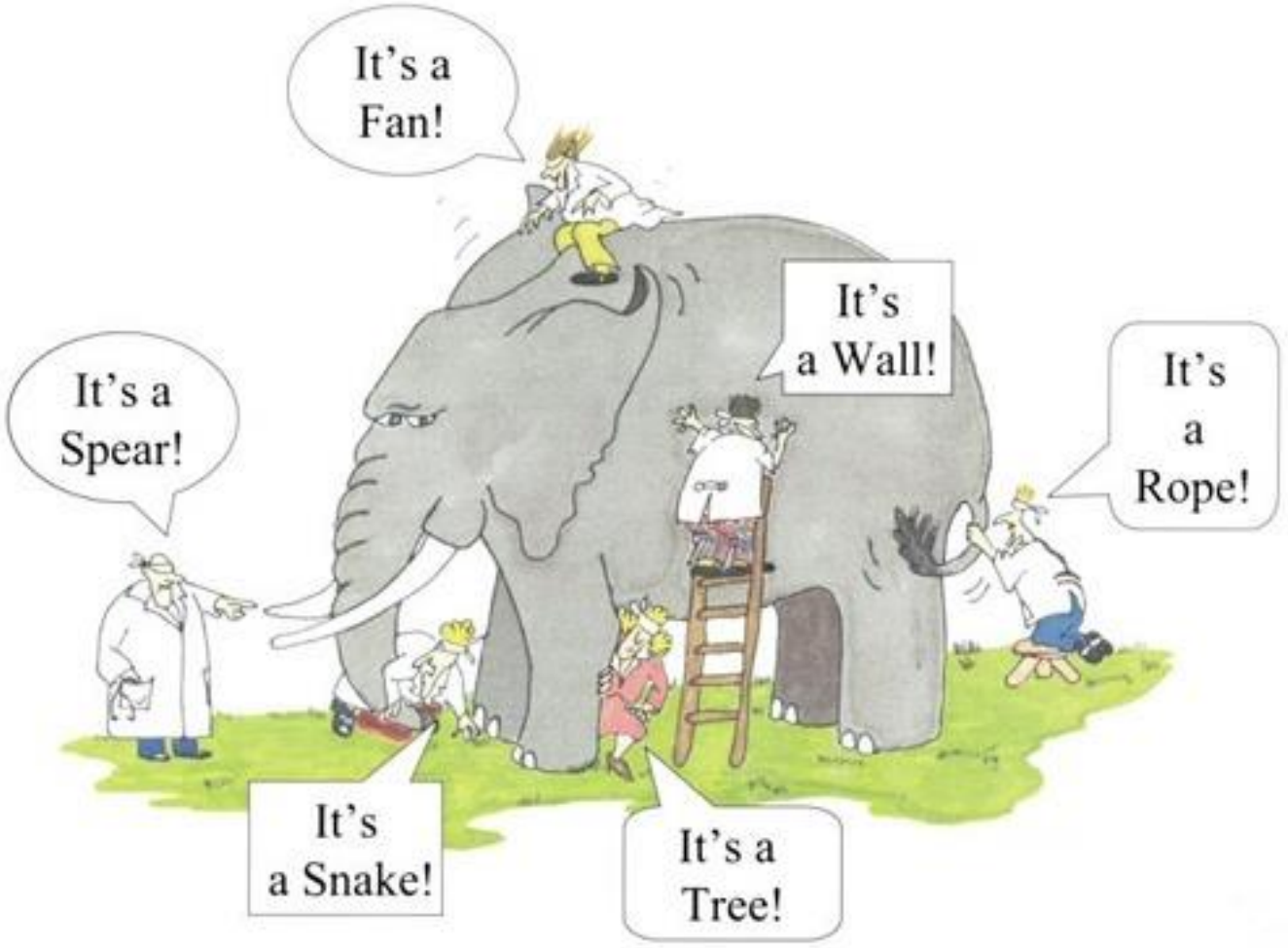
Summary 本章小结

- Machine learning is to study some algorithms that can learn from and make predictions on data.
  机器学习是研究一些可以从数据中学习、并对数据进行预测的算法。
- The different perspectives are aimed to try to have a taxonomy on the algorithms of machine learning, for being easy to understand machine learning.
  几个不同视角旨在尝试对机器学习的算法进行分类，以便于理解机器学习。
- Three perspectives on machine learning are proposed in this chapter, those are learning tasks, learning paradigms and learning models.
  本章提出了机器学习的三个视角，他们是：学习任务、学习范例以及学习模型。

The Three Perspectives 三个视角



The Three Perspectives 三个视角



Maybe "Blind Men and an Elephant"

Thank you for your attention!

AI