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.

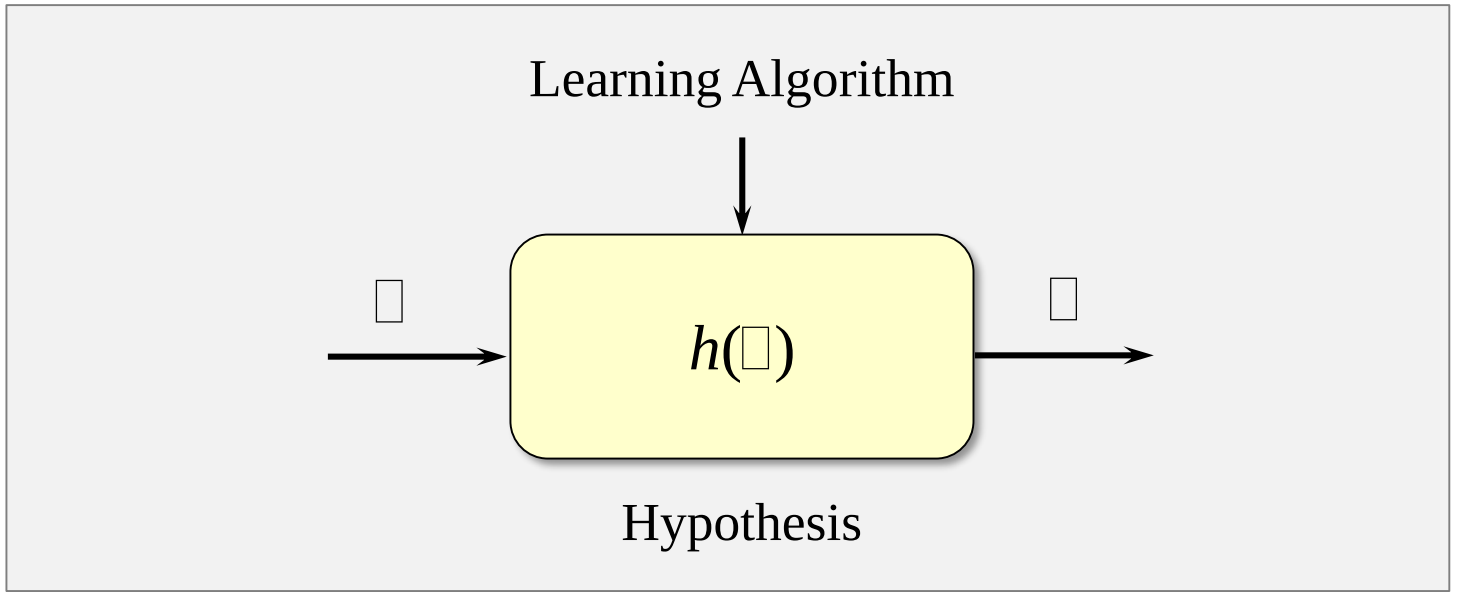
  没有训练过程。

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

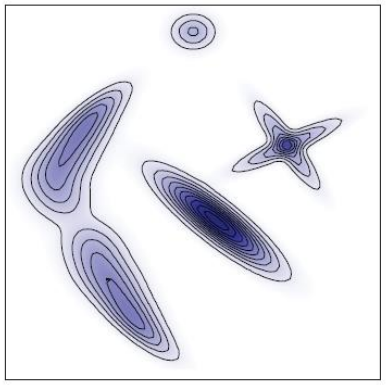
弗吉尼亚城市树冠测绘图中的分类树冠层次

11.2. Unsupervised Learning Paradigm

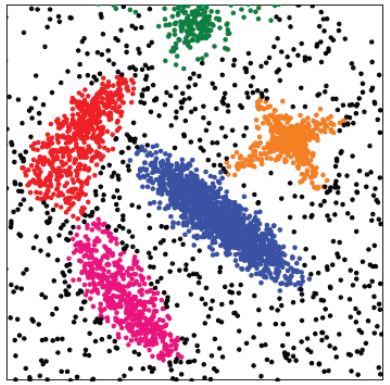
Contents:

Typical Clustering Algorithms 典型的聚类算法

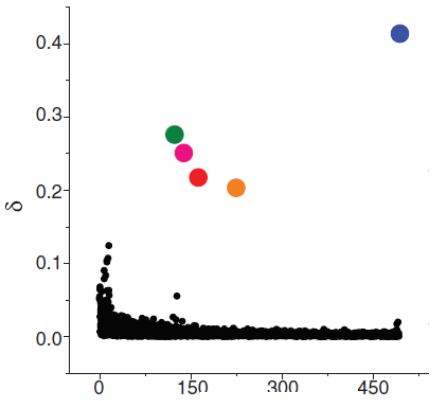
- Single-linkage clustering 单链聚类
- Conceptual clustering 概念聚类
- *k*-means *k*均值
- Fuzzy clustering 模糊聚类
- Clustering by density peaks 密度峰值聚类



(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。

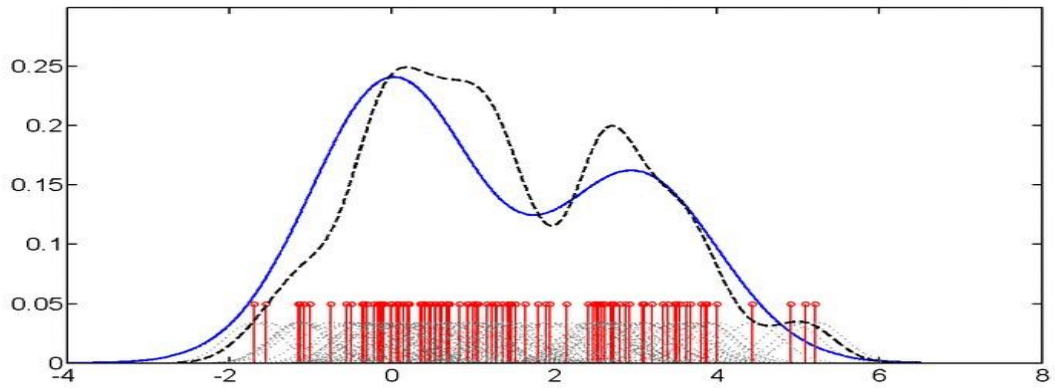
(c) The corresponding decision graph, with the centers colored by cluster.
对应的决策图，彩色点是聚类的中心点。

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

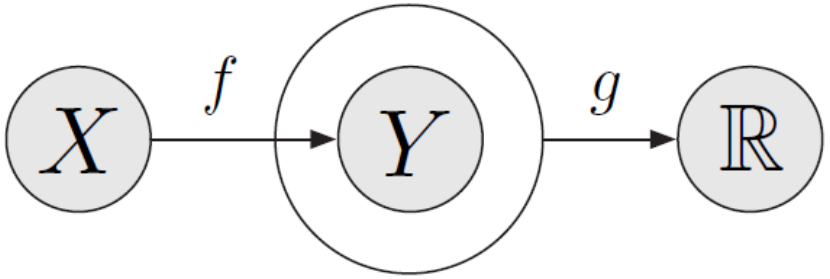
- Principal component analysis (PCA) 主成分分析 (PCA)
- Kernel PCA 核 PCA
- Graph-based kernel PCA 基于图的核 PCA
- Linear discriminant analysis (LDA) 线性判别分析 (LDA)
- Generalized discriminant analysis (GDA) 广义判别分析 (GDA)
- Multi-dimensional Scaling (MDS) 多维尺度分析 (MDS)
- Isometric feature mapping (Isomap) 等距特征映射 (Isomap)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

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

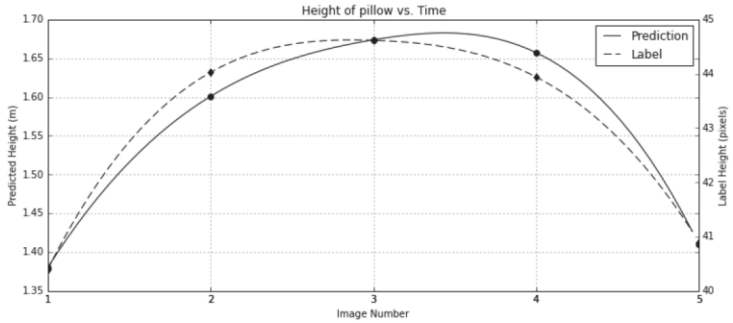
Case Study: Label-Free Supervision of Neural Networks 无标注有监督神经网络

- 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

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

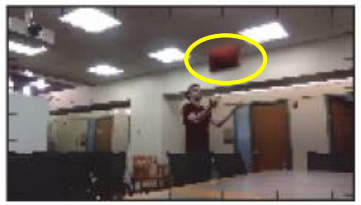
  - Constraint loss: 约束损失

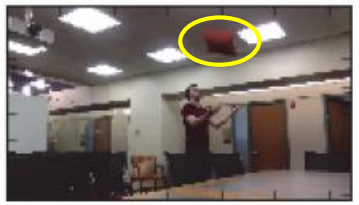
$$g(\mathbf{x}, f(\mathbf{x})) = g(f(\mathbf{x})) = \sum_{i=1}^{N} |\hat{\mathbf{y}}_i - f(\mathbf{x})_i|$$



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.

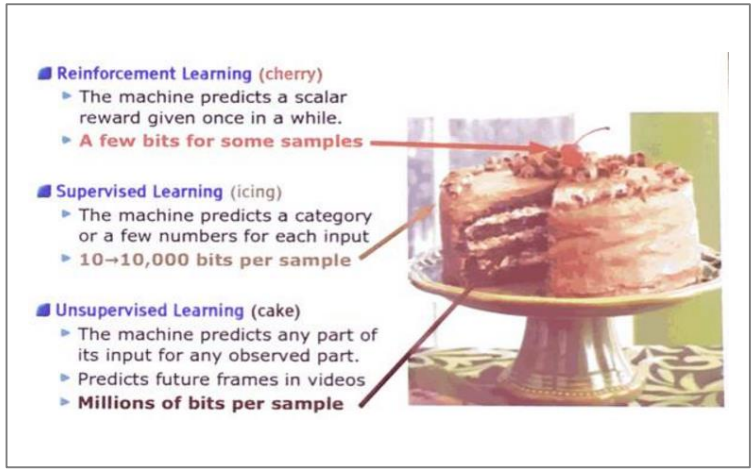
  如果智能是一块蛋糕，无监督学习就是这块蛋糕，有监督学习则是蛋糕上的糖霜，而强化学习是蛋糕上的樱桃。

- 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