



数据分析算法

北京理工大学计算机学院 孙新

2019年1月

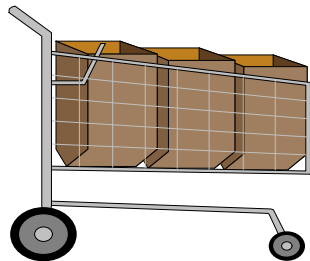
3.2 机器学习的典型任务——关联分析

- 关联分析：在交易数据、关系数据或其他信息载体中，查找存在于项目集合或对象集合之间的频繁模式、关联、相关性、或因果结构，简要描述就是挖掘在同一个数据集合中不同条目同时发生的概率
 - 有趣的故事“尿布与啤酒”：这是沃尔玛的真实案例，沃尔玛为了能够准确了解顾客在其门店的购买习惯，沃尔玛对其顾客的购物行为进行购物篮分析
 - 在原始交易数据的基础上进行分析和挖掘，发现背后的美国人的行为模式是：在美国，一些年轻的父亲下班后经常要到超市去买婴儿尿布，而他们中有30%~40%的人同时也为自己买一些啤酒

3.2 机器学习的典型任务——关联分析

定义：自然界中某种事物发生时其他事物也会发生，则这种联系称之为关联。反映事件之间依赖或关联的知识称为关联型知识（又称依赖关系）。要求找出描述这种关联的规则,并用以预测或识别。

目的是找出数据集中隐藏的关联，是离散变量因果分析的基础。

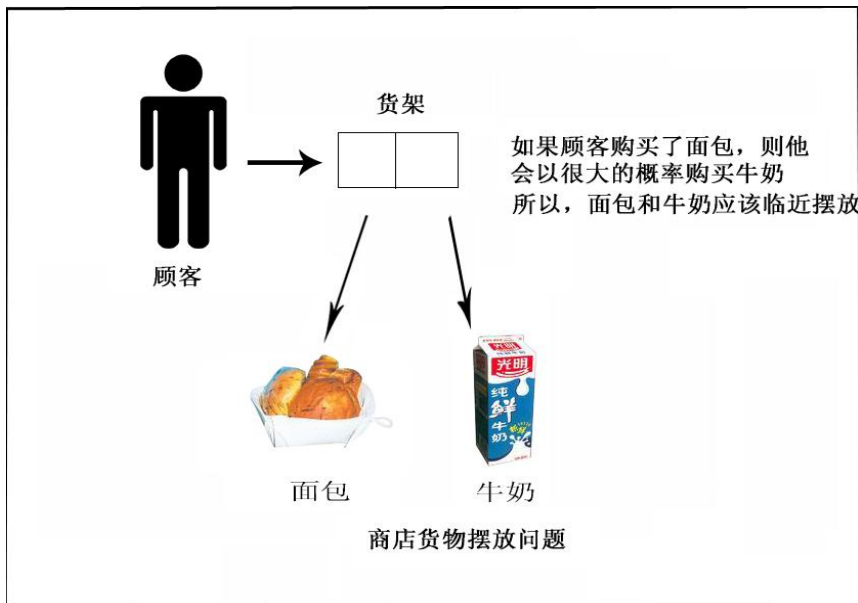


- 市场组合分析
- 套装产品分析
- 目录设计
- 交叉销售

3.2 机器学习的典型任务——关联分析

举例：通过发现顾客放入其购物篮中不同商品之间的联系，分析顾客的购买习惯。

这种信息可以引导销售，可以帮助零售商有选择地经销和安排货架。例如，将牛奶和面包尽可能放近一些，可以进一步刺激一次去商店同时购买这些商品。



关联分析 Association

3.2 机器学习的典型任务—小结

- ❑ 分类：将输入划分成两个或多个类别，通过一个给定的输入预测一个输出，这里的输出变量以类别形式展示
- ❑ 回归：通过一个给定的输入预测一个输出，这里的输出变量以实数的形式展示，是连续值而不是离散的。
- ❑ 聚类：输入被划分为若干个事先未知的组
- ❑ 关联：挖掘在同一个数据集合中不同条目同时发生的概率。

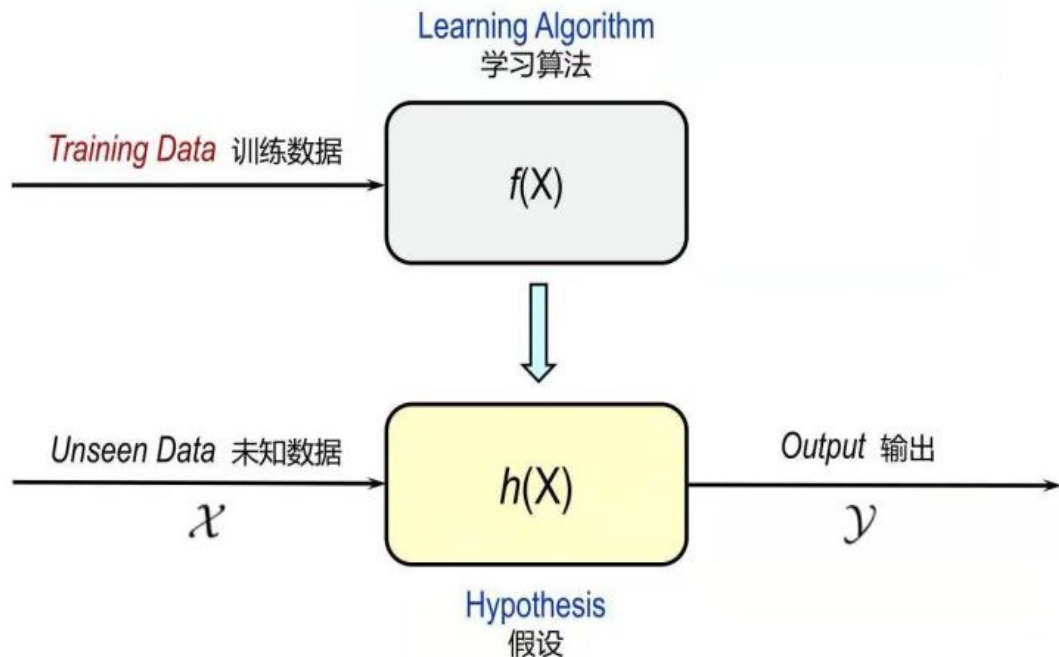
3.3 有监督学习和无监督学习

▣ 机器学习是如何工作的？

Supervised 有监督	算法采用一组标注好的数据进行训练，再对所有的未知点做出预测。
Unsupervised 无监督	算法仅接收未标注的数据，再对所有的未知点做出预测。
Reinforcement 强化	算法与外部环境交互，每个动作得到一个回报。

机器学习是如何工作的

- 机器学习算法就是在没有人类干预的情况下，从数据中学习，并在经验中改善的一种方法，学习任务可能包括学习从输入映射到输出的函数
- 可以这样来描述有监督学习：使用有标签的训练数据去学习从输入变量 (X) 到输出变量 (Y) 的映射函数。



3.3.1 有监督学习

- ❑ 机器学习算法接受一组标注的样本作为训练数据，然后对所有未知数据进行推测
- ❑ 算法生成从输入到输出的函数或者映射，然后可以将其用于对预先位置的数据生成输出

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

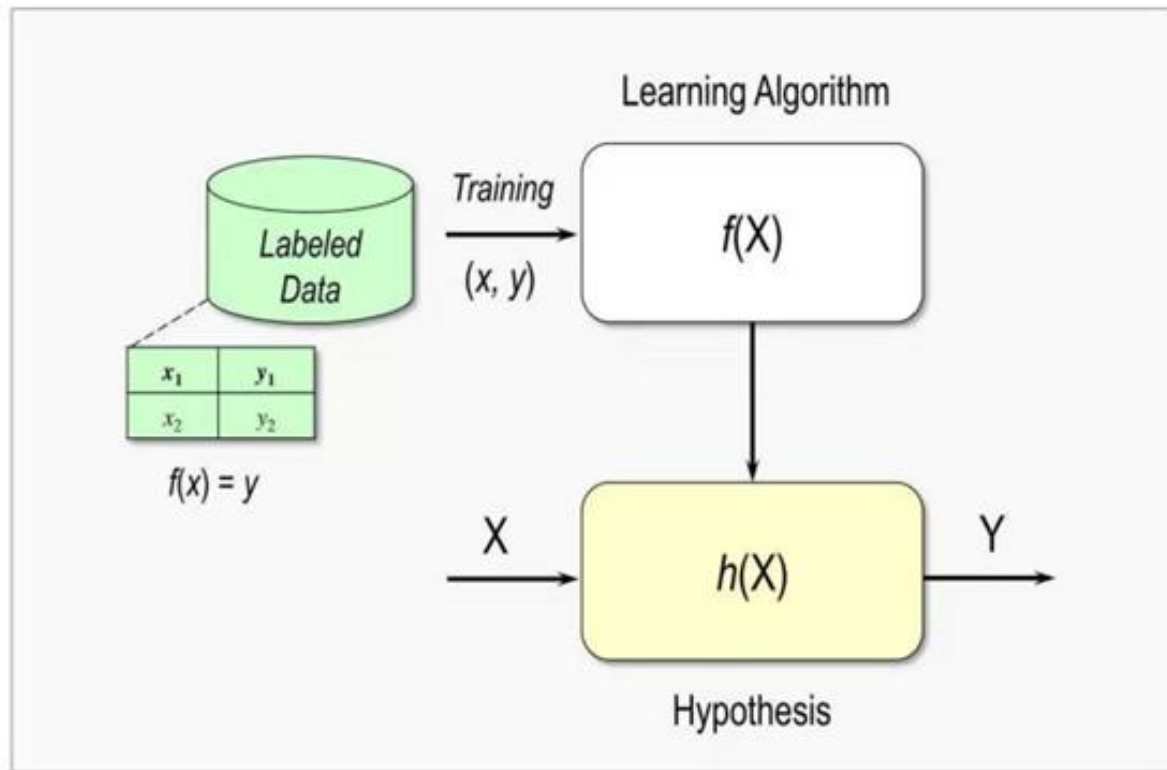
这是一种“教”学习算法的方式，就像“老师”讲授课程那样。

3.3.1 有监督学习

- 有监督学习中的训练数据是：
 - 每个训练数据具有一个已知标注，标注是由输入对象和预期输出值组成。例如：垃圾邮件与非垃圾邮件
- 训练后的模型（函数）可用于映射新的未知数据

3.3.1有监督学习

与有监督学习相关的任务



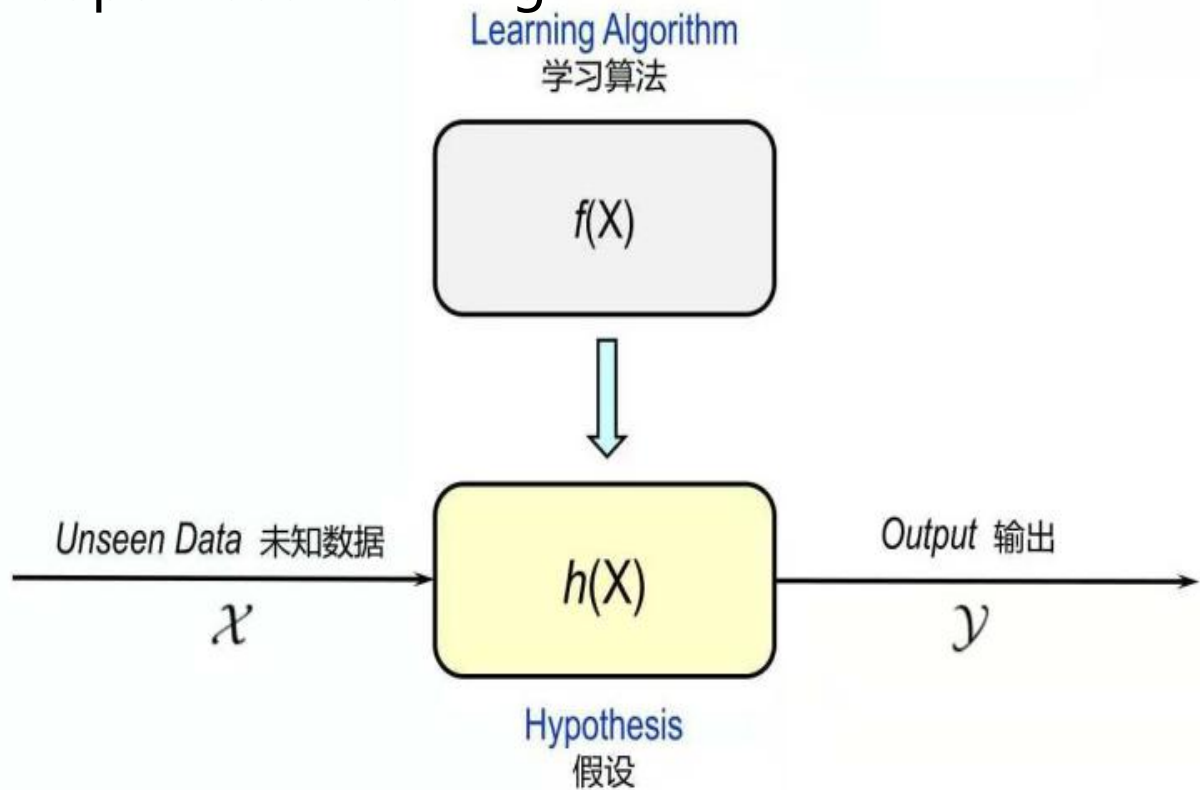
- ☐ Classification,
分类
- ☐ Regression,
回归
- ☐ Ranking.
排名

3.3.1 有监督学习

- 有监督学习示例
 - 垃圾邮件分类
 - 手写数字识别
 - 二手车价格预测

机器学习是如何工作的

□ 无监督学习 Unsupervised Learning



3.3.2 无监督学习

- ❑ 无监督学习就是指算法专门接收未标注数据，并对所有的未知点做出预测，
- ❑ 主要目标是发现数据中共性的东西，对观察值进行分类或者区分等

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

这是一种“自学”的方式，没有“老师”。

有监督与无监督学习

Supervised learning

有监督学习

- 给予学习器的样本是已标注的，
- 样本用于训练该算法。

Unsupervised learning

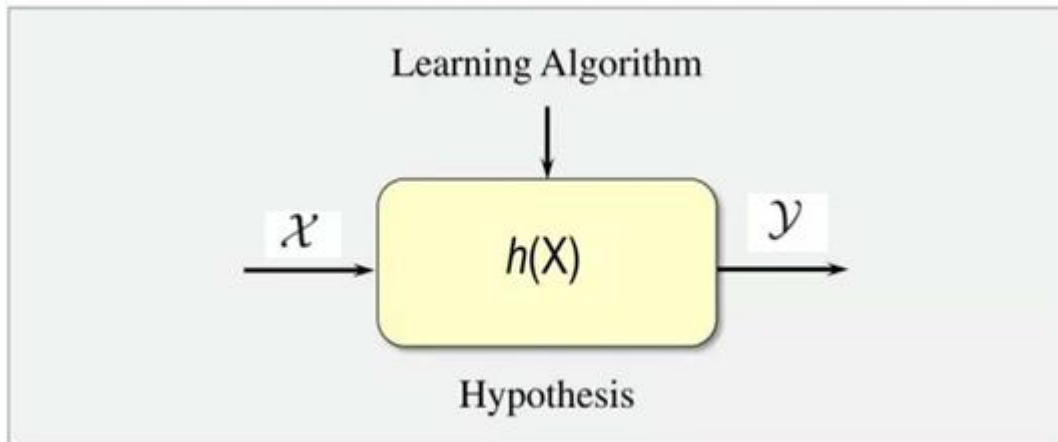
无监督学习

- 给予学习器的样本是未标注的，
- 没有训练过程。

3.3.2 无监督学习

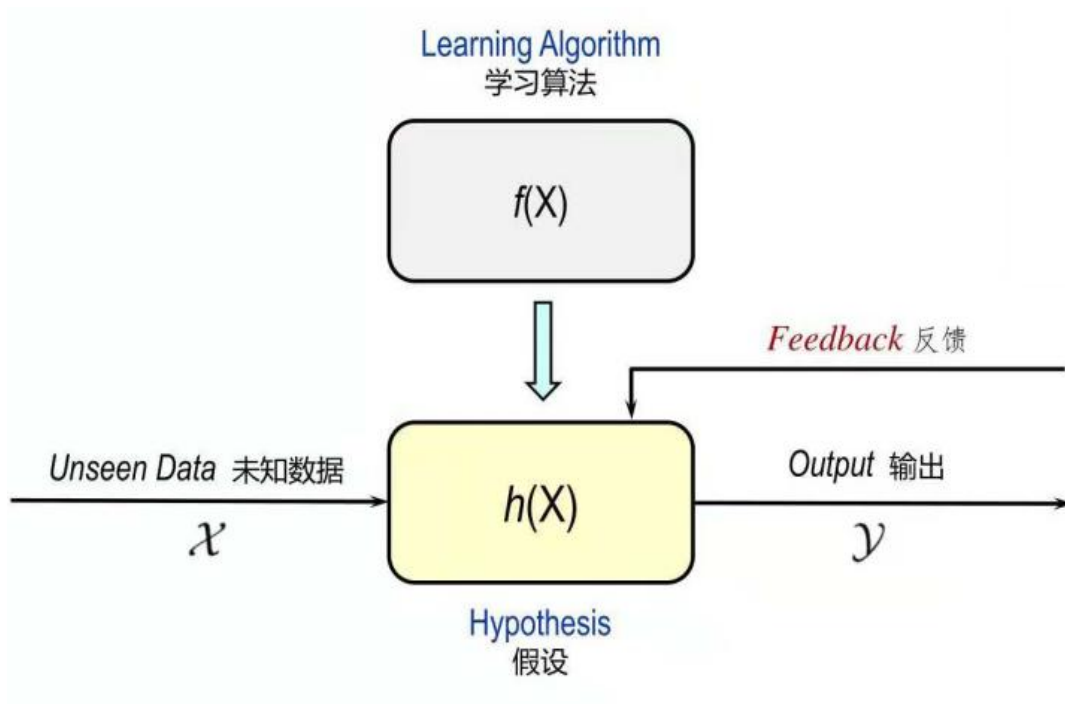
与无监督学习相关的任务

- ☐ 聚类
- ☐ 密度估计
- ☐ 降维



机器学习是如何工作的

强化学习 Reinforcement learning



小结

- ❑ 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。
- ❑ 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
 - ❑ 有监督学习和无监督学习
 - ❑ 机器学习的典型任务
 - ❑ 分类、回归 、聚类 、关联

谢 谢

Thank you for your attention!