

机器学习模型种类

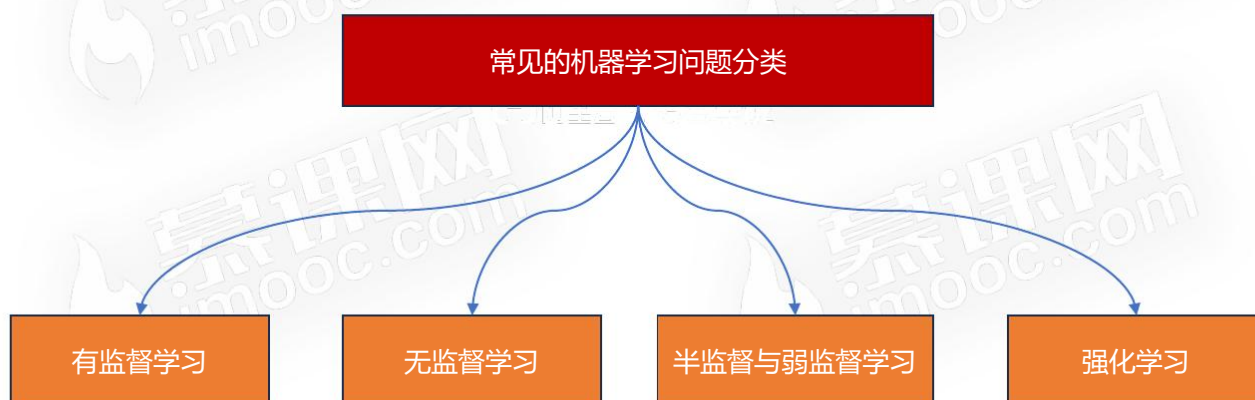
目录

- ◆ 机器学习问题概览
- ◆ 有监督学习模型
- ◆ 无监督学习模型

机器学习问题概览

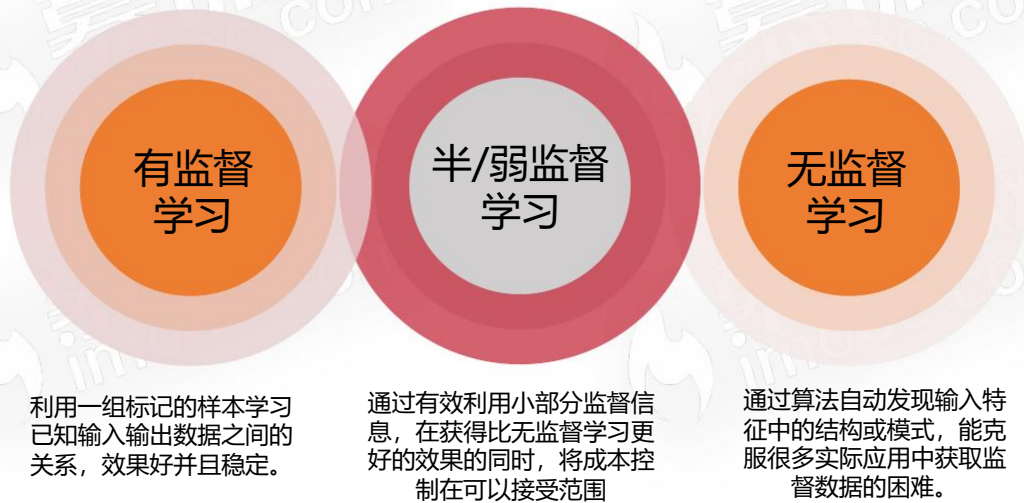
机器学习问题概览

◆ 机器学习问题分类



机器学习问题概览

◆ 监督学习、半监督学习、无监督学习的比较

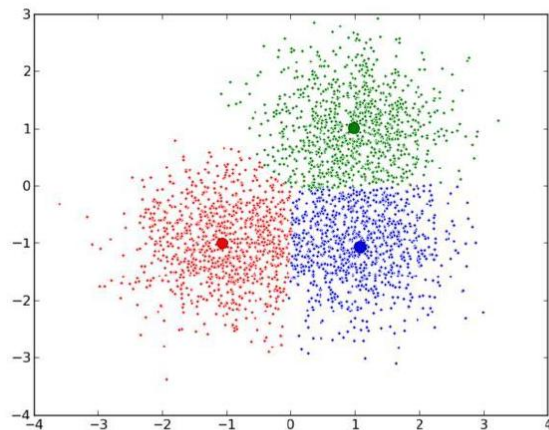


有监督学习与无监督学习

◆ 最常见的两类问题，根据是否存在真实的答案，即标签进行区分



有监督：在真值标签（分类类别）的指导下进行学习



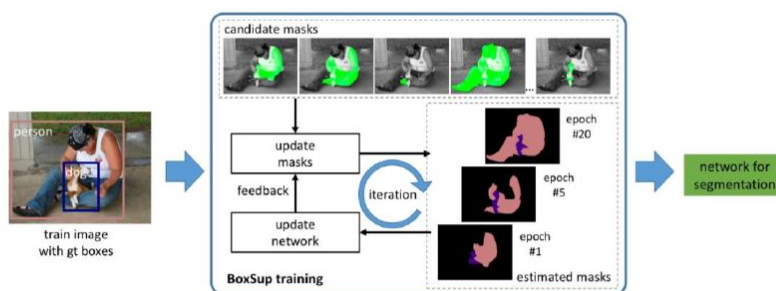
无监督：没有真值标签，学习数据的统计规律或潜在结构

半监督学习与弱监督学习

- ◆ 半监督即只有部分训练样本有标签值；弱监督即只有不完整的标签值



半监督分类任务，少量有标签数据与大量无标签数据



弱监督图像分割任务，标注目标框

强化学习

- ◆ 又称再励学习、评价学习或增强学习，在与环境的交互过程中学习策略，达成回报最大化或实现特定目标。



- 仅有奖励信号，没有监督
- 非瞬时的延迟反馈

- ◆ 训练数据和测试数据混合使用
- ◆ 不断与环境交互，通过试错，完成策略评估

有监督学习模型

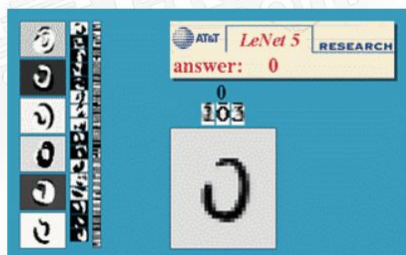
有监督学习模型

◆ 经典的有监督机器学习模型包括分类与回归模型

| 分类 | 回归 |
|-------------|-------------|
| 线性判别分析(LDA) | 线性回归 |
| 逻辑回归 | 局部加权回归 |
| 贝叶斯分类器 | 套索回归 |
| K近邻 | K近邻回归 |
| 感知器 | 弹性网络回归 |
| 决策树 | 回归树 |
| 随机森林 | 岭回归 |
| 支撑向量机(SVM) | 支撑向量回归(SVR) |
| Adaboost | 多项式回归 |

分类任务

- ◆ 将物理或抽象对象的集合分成多个不同的指定类别，标签为离散类别数值



图像分类



新闻分类



图像分割

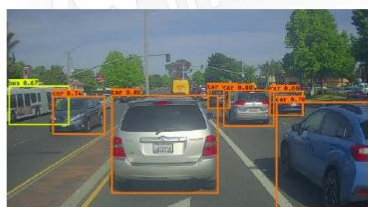
分词词性

| | | | | | | | | | | |
|----------|------|------------|-----|------------|-----|------|-----|-----|------|--------|
| 谢尔盖·列昂诺夫 | { | 1907年1月12日 | ~ | 1966年1月14日 | } | , | 原 | 苏联 | 宇航 | |
| PPS | W | DATE | W | DATE | W | W | LOC | LOC | N | |
| 事业 | 的 | 伟大 | 设计师 | 与 | 细照者 | , | 第一枚 | 射段 | 超过 | 8000公里 |
| N | U | A | N | C | N | W | M | N | V | M |
| 的 | 洲际 | 火箭 | { | 弹道导弹 | } | 的 | 设计者 | , | 第一颗 | 人造地球卫星 |
| U | N | N | W | N | W | U | N | W | N | N |
| 的 | 运载火箭 | 的 | 设计者 | , | 第一颗 | 载人航天 | 飞船 | 的 | 总设计师 | , |
| U | N | U | N | W | M | N | N | U | N | W |

词性标注

回归任务

- ◆ 将输入的数据拟合成连续的输出结果，标签为连续数值



边框回归



关键点回归



时序数值预测

无监督学习模型

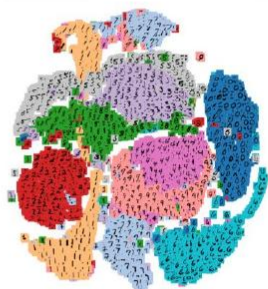
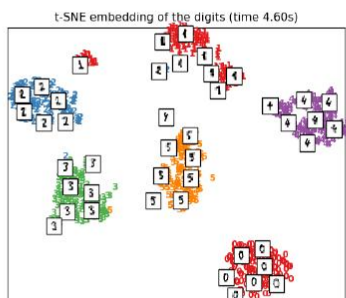
无监督学习模型

◆ 经典的无监督机器学习模型包括降维、聚类与生成模型

| 降维 | 聚类 | 生成 |
|----------|-----------------------------|--------------|
| PCA | Kmeans | AE/VAE |
| 核PCA | Fuzzy C-means | GAN |
| LDA | 层次聚类Hierarchical Clustering | 扩散模型 |
| 局部线性嵌入 | DBSCAN | 流模型 |
| SNE与tSNE | 谱聚类(Spectral Clustering) | GPT3/ChatGPT |

降维任务

- ◆ 采用某种映射方法，将原高维空间中的数据点映射到低维度空间中，对特征进行降维与可视化



降维后的数据更易于处理，且可以可视化（人最多能看到三维空间）

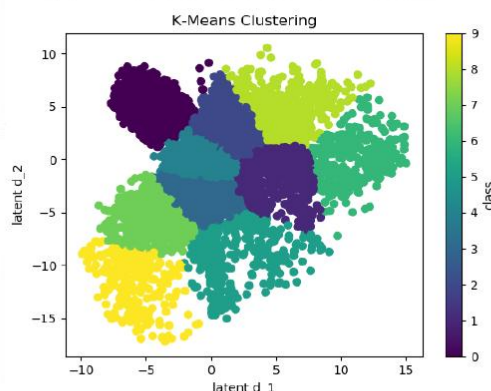
将0-9手写数字图像投影到2维空间的结果

聚类任务

- ◆ 将集合分成由类似的对象组成的多个类的过程，是分类问题，但没有训练过程



=>



把一批样本划分成多个类，使得在某种相似度指标下每一类中的样本尽量相似，不同类的样本之间尽量不同

生成任务

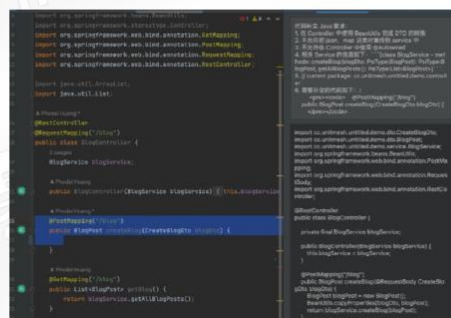
- ◆ 估计特定集合的概率密度函数，生成高质量文本、图像、语音、代码等数据



图像生成



文本生成



代码生成

下次预告：机器学习模型评估