

Class 0 机器学习基本概念

2018年11月16日 9:50

• 几个问题：

• Q：为什么最近几年人工智能再次火爆？

A： 1.计算能力的提升；
2.算法优化；
3.大数据时代的到来。

• Q：人工智能、机器学习以及深度学习是什么样的关系？

A：人工智能的领域范围最广，深度学习领域最窄。

机器学习是人工智能的子领域，而深度学习是机器学习的一个分支。



• 机器学习的基本概念：

• 机器学习的定义：

“无需明确编程而给予计算机学习能力的研究领域。”（1959）

“如果它在任务T中的性能被P测试，随着经验E的增加而提高，则被认为是计算机程序从经验E中学习某些类别的任务T和性能度量P。”（1997）

• 机器学习的特点：

1. 机器学习算法以数据和特征为基础，是数据驱动的科学；

2. 机器学习的目标是对数据进行预测与分析;
3. 机器学习以模型方法为中心, 利用统计学习的方法构建模型, 并且利用构建好的模型对未知的数据进行预测和分类;
4. 机器学习是以概率论、统计学、计算理论以及计算机科学等多领域交叉的学科。

- 机器学习的分类:

1. 监督学习:

训练数据既包含特征向量又包含类别标签信息。

- 分类、回归等

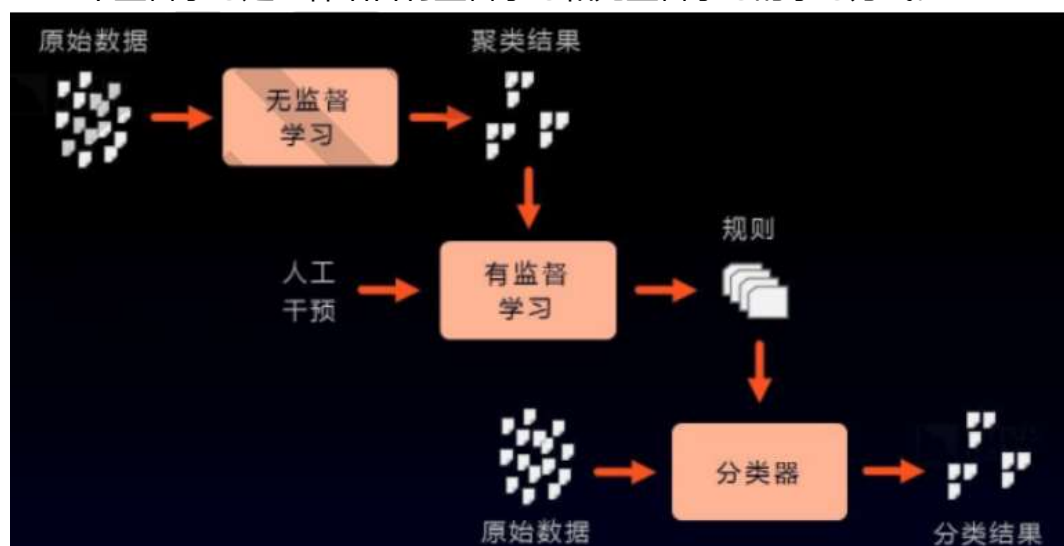
2. 无监督学习:

训练数据只有特征向量而没有类别标签信息。

- 聚类

3. 半监督学习:

半监督学习是一种结合有监督学习和无监督学习的学习方式。



4. 强化学习、深度学习...

- 机器学习阶段:

- 训练阶段: 用训练数据训练模型;

- **测试阶段：**用测试数据评估模型；
- **应用阶段：**用将来的数据预测结果。

- **机器学习过程：**

- **假设：**
 - 具有（未知）参数（或结构）的数学模型
- **学习：**
 - 用来评估参数
 - 最大似然估计（MLE），MAP，贝叶斯估计
 - 代价函数优化
- **决策：**
 - 贝叶斯决策
 - 预测函数

- **比较：**

- **分类 vs. 回归：**
 - 回归：输出变量为连续变量的预测问题是回归问题；
 - 线性回归
 - 分类：输出变量为离散变量的预测问题是分类问题；
 - 逻辑回归、Softmax回归、感知机
- **生成式模型 vs. 判别式模型：**
 - 判别式模型：直接对后验概率 $p(y|x;\theta)$ 建模；
 - 线性回归、逻辑回归、Softmax回归、支持向量机SVM、神经网络
 - 生成式模型：对x和y的联合分布 $p(x,y)$ 建模，通过贝叶斯公式来求得 $p(y_i|x)$ ；
 - 朴素贝叶斯模型

- 准则函数：

$$\min \left\{ L(\theta) = \sum (\hat{y} - y)^2 \right\}$$

- 最小均方误差 (LMS) : $\min \left\{ L(\theta) = - \sum y * \log(\hat{y}) \right\}$

- 最小交叉熵 (CE) : $\max \left\{ L(\theta) = \sum \log p(y|x; \theta) \right\}$

- 最大似然估计 (MLE) :

- 感知机准则

- 最优化方法：

- **梯度下降**：沿梯度下降的方向求解极小值；

- 批处理梯度下降 (GD) : 每次更新，需计算所有样本；
- 随机梯度下降 (SGD) : 每次更新，只需计算某一个样本；

- **牛顿法**：用函数f(x)的泰勒级数的前面几项来寻找方程f(x)=0的根；

- 与梯度下降法的比较：
 - 梯度下降法的迭代公式

$$\theta^{(t+1)} = \theta^{(t)} - \alpha \nabla J(\theta^{(t)})$$

- 牛顿法的迭代公式

$$\theta^{(t+1)} = \theta^{(t)} - H^{-1} \nabla J(\theta^{(t)})$$

与梯度下降法相比，牛顿法只是将学习率 α 替换成了 H^{-1} ；

这可以理解为：

牛顿法根据代价函数的二阶导数信息，自动计算出了合适的学习率；

因此有更快的迭代速度,而作为交换，牛顿法需要计算庞大的Hessian矩阵，矩阵的大小为 $m \times m$ ，计算速度慢，消耗资源大。