

三、Types of Learning 机器学习的分类

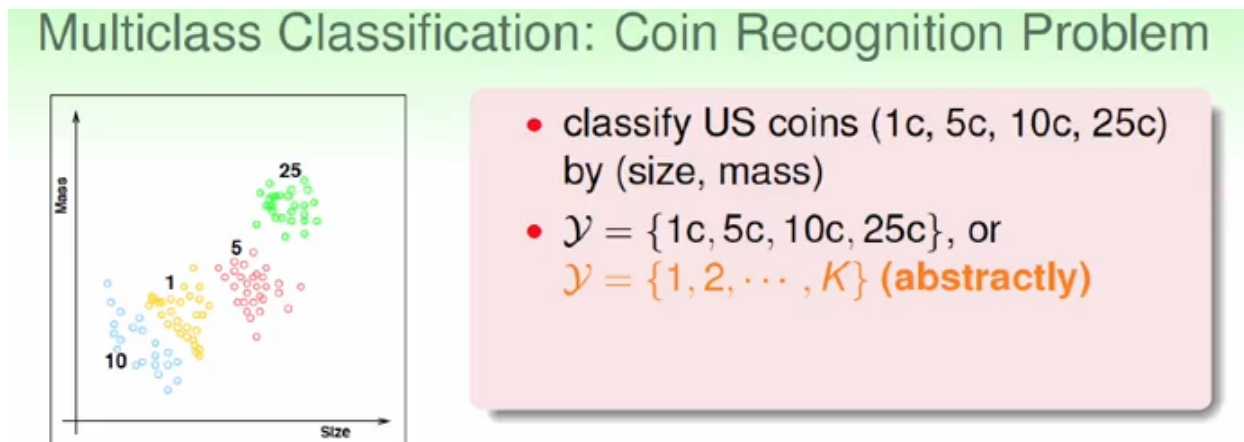
3.1 Learning with Different Output Space 不同类型的输出空间

3.1.1 是非题 -- binary classification 二元分类

为线性可分（linear binary separable），如可以使用PLA求解；b是包含噪音可以使用pocket求解，而c会在后面章节中详细叙述，属于多项式可分解。当然解决以上三种二元分类问题的机器学习方法很多，因为二元分类问题是机器学习中很重要、核心的问题。

3.1.2 Multiclass Classification 多元分类问题

1. 硬币分类，硬币共有4种。抽象的来说可以分类K中，K可以用1，2，3，4来表示。根据不同的硬币特征，画出图如下：



2. 二元分类是特殊的多元分类，即K=2的情况
3. 应用方面：
 - 手写识别数字 0-9
 - 一张图里有水果识别是哪一种
 - 分类电子邮件是重要的、社交的、广告的等
 - 尤其是视觉或听觉的辨析
4. 现在的输出Y通常不再是-1和1两种，而是一个范围输出，比如成绩打分范围是0-100，医院预测病人恢复需要多少天等等，这种在统计学里面做回归分析（Regression）

3.1.3 Regression 回归分析

1. 输出是一个实数

多类别分类问题，式其他很多复杂问题的延伸。例如，在自然语言处理方面，一个英文句子每个单词的词性是什么？这可以看做是一个更复杂的多类别问题，可以从多个句子中学的一个句子的结构。这是由多个单词共同组成的一个结构。这又叫做结构学习（structured learning）

3.3.4 结构学习

当然还有其他更为复杂的问题，比如很多很多类型的分类问题。

3.2 Learning with Different Data Label 不同的数据标记

监督式学习 supervised learning

样本数据有明确的标签，即答案。例如对于一个硬币分类问题，数据已经明确这个硬币是什么了。

非监督式学习 unsupervised learning

样本数据没有标签，不知道是什么

半监督式学习 semi-supervised learning

例如要对树分类，树的样本非常庞大，我们只能对一部分数据进行标注标签，而其他大多数数据是没有标签的，可以通过这些有标签的来帮助没有标签的学习

强化学习 reinforcement learning

是一种惩罚/奖励机制。例如我们训练狗狗时候，让狗狗坐下，狗狗却尿尿，我们惩罚它。狗狗跟我们握手，我们觉得虽然不是坐下，但也是一种好的行为，便奖励它。通过大量的训练，以达到好的结果。

active learning

举个例子，一个摄影师有很多照片，他想把这些照片分类，但是他没有时间一个一个去标注。于是写一个算法让机器自动分类，当机器觉得对某张照片的分类不是很有信心的时候便询问摄影师，让摄影师来决定，这“问”的过程，便是让机器主动问问题的过程，也就是active。

3.3 Learning with Different Protocol 不同方式获取数据

对此节的内容进行简单阐述，在不同的协议中可以将机器学习分为三大类：

1. 批量（batch）学习就是将很多数据一次性的给算法进行学习，最常见的方式；
2. 在线（online）学习就是一点一点将数据传输进去，如PLA和增强学习都适用于这种形式；
3. 主动（active）学习是主动提出问题让算法解决，可以节省大量的训练和标记消耗。

3.4 Learning with Different Input Space 不同的输入空间

输入又可以称之为特征（features），其主要分为三种：

1. 具体特征（Concrete Features），具体特征最大特点就是便于机器学习的处理，也是基础篇中主要讨论的情形。这种情况是人类或者机器通过一定的方式提取获得的，具有实用性。
2. 原始特征（Raw Features），如图片的像素等等，是最为常见到的资料，但是需要经过处理，转换成具体特征，才容易使用，实用性不太大。
3. 抽象特征（Abstract Features），如一些ID之类的看似无意义的数字，这就更需要特征的转换、提取等工作（相对于原始特征而言），几乎没有实用性。