机器学习主要分为两类：有监督的机器学习和无监督的机器学习。所谓有监督的机器学习，指的是：已知道训练集数据的类别（label），用来训练分类器（实质上是个分类task）；而无监督的机器学习，指的是：不知道训练集数据的类别（label），用来训练分类器（实质是一种聚类）。

分类任务的本质就是通过很多有标记的数据，训练处一个模型，然后利用该模型，对输入的，进行预测输出的。该模型一般有两种形式：

1. 判别函数；
2. 条件概率分布。

从训练分类模型的方法上讲，可以分为两类，一类是生成方法，一类是判别方法。判别函数模型采用的是典型的判别方法；而条件概率分布的确定可以分为生成方法和判别方法。

下面以一个二分类例子说明下如何用生成方法和判别方法做训练：

假设一组训练数据集个数据，第个数据用特征向量表示；这个数据分为、

两类别，假设类别的数据有个，类别的数据有个，。对于一个测试数据，我们目标是求出。

**概率生成模型：**

利用贝叶斯公式：

去



其中：



 代表从类中产生的概率。假设类中的个数据服从参数为和的高斯分布，那如何求出这个高斯分布的参数呢？用（最大似然法）求，其中：



为了方便计算，对似然函数取对数后求导，令导数等于零，可得出：







则测试数据来自于类的概率：



同理：



此时，再利用贝叶斯公式就可以求出后验分布函数：



当则分类为，否则分类为。

**概率判别模型：**

概率判别模型与生产模型的差异在于：判别模型不会去考虑测试数据与类或类数据的内在关系，不会去模拟用类或类概率分布来生成数据，而是只考虑测试数据本身的差异性，利用这种差异性来区分不同的类别。



假设这个测试数据基于后验概率分布产生，这里我们假设用函数作为激活函数，。

给定一组和，最有可能产生这组训练数据的和是多少？还是用（最大似然法）求解：







最后可得：



其中，：11，02。

现在的目标就是让 最小，采用梯度下降法求得。



**概率判别模型的解题思路：**

如上例所示，概率判别模型是通过训练数据得到一个判别式，实际上该式是由训练数据经过训练后得到的：，同时我们认为这个后验概率分布是由参数决定，。

问题转变为如何推：

？

实际上：



问题再次转变为：求条件分布参数关于训练数据的后验概率分布；

利用贝叶斯公式可得：





由于这两个积分的求解太麻烦，通过方法：如果训练样本足够多，则可以使用的最大后验分布对参数进行点估计。



如果在似然函数较大区间内是固定不变或者变化较小的话，问题又转化为求最大似然函数。

**生成方法特点：**

生成方法学习的是联合概率密度分布，可以从统计的角度表示分布的情况，能够反映同类数据本身的相似度，它不关心到底划分不同类的分类边界在哪。生成方法的学习收敛速度更快，当样本容量增加时，学习到的模型可以更快的收敛到真实模型，当存在隐变量时，依旧可以用生成方法学习，此时判别方法就不行。

**判别方法特点：**

直接学习决策函数或者条件概率，不能反映训练数据本身的特性，但它寻找不同类别之间的最优分裂面，反映的是异类数据之间的差异，直接面对预测往往学习准确度更高。

简而言之，生成模型是从大量数据中找规律，属于统计学习；判别模型只关心不同类型的数据的差异，利用差别来区分类别。

  

2017.07.22