生成模型与判别模型

一直在看论文的过程中遇到这个问题，折腾了不少时间，然后是下面的一点理解，不知道正确否。若有错误，还望各位前辈不吝指正，以免小弟一错再错。在此谢过。

一、决策函数 Y=f(X)或者条件概率分布 P(Y|X)

监督学习的任务就是从数据中学习一个模型（也叫分类器），应用这一模型，对给定的输入 X 预测相应的输出 Y。这个模型的一般形式为决策函数 Y=f(X)或者条件概率分布 P(Y|X)。

决策函数 Y=f(X)：你输入一个 X，它就输出一个 Y，这个 Y 与一个阈值比较，根据比较结果判定 X 属于哪个类别。例如两类（w1 和 w2）分类问题，如果 Y 大于阈值，X 就属于类 w1，如果小于阈值就属于类 w2。这样就得到了该 X 对应的类别了。

条件概率分布 P(Y|X)：你输入一个 X，它通过比较它属于所有类的概率，然后输出概率最大的那个作为该 X 对应的类别。例如：如果 P(w1|X)大于 P(w2|X)，那么我们就认为 X 是属于 w1 类的。

所以上面两个模型都可以实现对给定的输入 X 预测相应的输出 Y 的功能。实际上通过条件概率分布 P(Y|X)进行预测也是隐含着表达成决策函数 Y=f(X)的形式的。例如也是两类 w1 和 w2，那么我们求得了 P(w1|X)和 P(w2|X)，那么实际上判别函数就可以表示为 Y=

P(w1|X)/P(w2|X)，如果 Y 大于 1 或者某个阈值，那么 X 就属于类 w1，如果小于阈值就属于类 w2。而同样，很神奇的一件事是，实际上决策函数 Y=f(X)也是隐含着使用 P(Y|X)的。因为一般决策函数 Y=f(X)是通过学习算法使你的预测和训练数据之间的误差平方最小化，而贝叶斯告诉我们，虽然它没有显式的运用贝叶斯或者以某种形式计算概率，但它实际上也是在隐含的输出极大似然假设（MAP 假设）。也就是说学习器的任务是在所有假设模型有相等的先验概率条件下，输出极大似然假设。

所以呢，分类器的设计就是在给定训练数据的基础上估计其概率模型 P(Y|X)。如果可以估计出来，那么就可以分类了。但是一般来说，概率模型是比较难估计的。给一堆数给你，特别是数不多的时候，你一般很难找到这些数满足什么规律吧。那能否不依赖概率模型直接设计分类器呢？事实上，分类器就是一个决策函数（或决策面），如果能够从要解决的问题和训练样本出发直接求出判别函数，就不用估计概率模型了，这就是决策函数 Y=f(X)的伟大使命了。例如支持向量机，我已经知道它的决策函数（分类面）是线性的了，也就是可以表示成 Y=f(X)=WX+b 的形式，那么我们通过训练样本来学习得到 W 和 b 的值就可以得到 Y=f(X)了。还有一种更直接的分类方法，它不用事先设计分类器，而是只确定分类原则，根据已知样本（训练样本）直接对未知样本进行分类。包括近邻法，它不会在进行具体的预测之前求出概率模型 P(Y|X)或者决策函数 Y=f(X)，而是在真正预测的时候，将 X 与训练数据的各类的 Xi 比较，和哪些比较相似，就判断它 X 也属于 Xi 对应的类。

实际上，说了那么多，也不知道自己表达清楚了没有。那我们是谈生成模型和判别模型，上面到底啰嗦了那么多到底有啥阴谋啊？呵呵，往下说就知道了。

二、生成方法和判别方法

监督学习方法又分生成方法（Generative approach）和判别方法（Discriminative approach），所学到的模型分别称为生成模型（Generative Model）和判别模型（Discriminative Model）。咱们先谈判别方法，因为它和前面说的都差不多，比较容易明白。

判别方法：由数据直接学习决策函数 Y=f(X)或者条件概率分布 P(Y|X)作为预测的模型，即判别模型。基本思想是有限样本条件下建立判别函数，不考虑样本的产生模型，直接研究预测模型。典型的判别模型包括 k 近邻，感知级，决策树，支持向量机等。

生成方法：由数据学习联合概率密度分布 P(X,Y)，然后求出条件概率分布 P(Y|X)作为预测的模型，即生成模型：P(Y|X)= P(X,Y)/ P(X)。基本思想是首先建立样本的联合概率概率密度模型 P(X,Y)，然后再得到后验概率 P(Y|X)，再利用它进行分类，就像上面说的那样。注意了哦，这里是先求出 P(X,Y)才得到 P(Y|X)的，然后这个过程还得先求出 P(X)。P(X)就是你的训练数据的概率分布。哎，刚才说了，需要你的数据样本非常多的时候，你得到的 P(X)才能很好的描述你数据真正的分布。例如你投硬币，你试了 100 次，得到正面的次数和你的试验次数的比可能是

3/10，然后你直觉告诉你，可能不对，然后你再试了 500 次，哎，这次正面的次数和你的试验次数的比可能就变成 4/10，这时候你半信半疑，不相信上帝还有一个手，所以你再试 200000 次，这时候正面的次数和你的试验次数的比（就可以当成是正面的概率了）就变成 5/10 了。这时候，你就觉得很靠谱了，觉得自己就是那个上帝了。呵呵，真啰嗦，还差点离题了。

还有一个问题就是，在机器学习领域有个约定俗成的说法是：不要去学那些对这个任务没用的东西。例如，对于一个分类任务：对一个给定的输入 x，将它划分到一个类 y 中。那么，如果我们用生成模型：p(x,y)=p(y|x).p(x)

那么，我们就需要去对 p(x)建模，但这增加了我们的工作量，这让我们很不爽（除了上面说的那个估计得到 P(X)可能不太准确外）。实际上，因为数据的稀疏性，导致我们都是被强迫地使用弱独立性假设去对 p(x)建模的，所以就产生了局限性。所以我们更趋向于直观的使用判别模型去分类。

这样的方法之所以称为生成方法，是因为模型表示了给定输入 X 产生输出 Y 的生成关系。用于随机生成的观察值建模，特别是在给定某些隐藏参数情况下。典型的生成模型有：朴素贝叶斯和隐马尔科夫模型等。

三、生成模型和判别模型的优缺点

在监督学习中，两种方法各有优缺点，适合于不同条件的学习问题。

生成方法的特点：上面说到，生成方法学习联合概率密度分布P(X,Y)，所以就可以从统计的角度表示数据的分布情况，能够反映同类数据本身的相似度。但它不关心到底划分各类的那个分类边界在哪。生成方法可以还原出联合概率分布 P(Y|X)，而判别方法不能。生成方法的学习收敛速度更快，即当样本容量增加的时候，学到的模型可以更快的收敛于真实模型，当存在隐变量时，仍可以用生成方法学习。此时判别方法就不能用。

判别方法的特点：判别方法直接学习的是决策函数 Y=f(X)或者条件概率分布 P(Y|X)。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类面，反映的是异类数据之间的差异。直接面对预测，往往学习的准确率更高。由于直接学习 P(Y|X)或 P(X)，可以对数据进行各种程度上的抽象、定义特征并使用特征，因此可以简化学习问题。

四、生成模型和判别模型的联系

由生成模型可以得到判别模型，但由判别模型得不到生成模型。

五、再形象点可以吗

例如我们有一个输入数据 x，然后我们想将它分类为标签 y。（迎面走过来一个人，你告诉我这个是男的还是女的）

生成模型学习联合概率分布 p(x,y)，而判别模型学习条件概率分布 p(y|x)。

下面是个简单的例子：

例如我们有以下(x,y)形式的数据：(1,0), (1,0), (2,0), (2, 1)

那么 p(x,y)是：

	y=0	y=1
x=1	1/2	0
x=2	1/4	1/4

而 p(y|x) 是：

	y=0	y=1
x=1	1	0
x=2	1/2	1/2

我们为了将一个样本 x 分类到一个类 y，最自然的做法就是条件概率分布 p(y|x)，这就是为什么我们对其直接求 p(y|x)方法叫做判别算法。而生成算法求 p(x,y)，而 p(x,y)可以通过贝叶斯方法转化为 p(y|x)，然后再用其分类。但是 p(x,y)还有其他作用，例如，你可以用它去生成(x,y)对。

再假如你的任务是识别一个语音属于哪种语言。例如对面一个人走过来，和你说了一句话，你需要识别出她说的到底是汉语、英语还是法语等。那么你可以有两种方法达到这个目的：

1、学习每一种语言，你花了大量精力把汉语、英语和法语等都学会了，我指的学会是你知道什么样的语音对应什么样的语言。然后再有人过来对你哄，你就可以知道他说的是什么语音，你就可以骂他是“米国人还是小日本了”。（呵呵，切勿将政治掺杂在技术里面）

2、不去学习每一种语言，你只学习这些语言模型之间的差别，然后再分类。意思是指我学会了汉语和英语等语言的发音是有差别的，我学会这种差别就好了。

那么第一种方法就是生成方法，第二种方法是判别方法。

生成算法尝试去找到底这个数据是怎么生成的（产生的），然后再对一个信号进行分类。基于你的生成假设，那么那个类别最有可能产生这个信号，这个信号就属于那个类别。判别模型不关心数据是怎么生

成的，它只关心信号之间的差别，然后用差别来简单对给定的一个信号进行分类。

六、对于跟踪算法

跟踪算法一般来说可以分为两类：基于外观模型的生成模型或者基于外观模型的判别模型。

生成模型：一般是学习一个代表目标的模型，然后通过它去搜索图像区域，然后最小化重构误差。类似于生成模型描述一个目标，然后就是模式匹配了，在图像中找到和这个模型最匹配的区域，就是目标了。

判别模型：将跟踪问题看成一个二分类问题，然后找到目标和背景的决策边界。它不管目标是怎么描述的，那只要知道目标和背景的差别在哪，然后你给一个图像，它看它处于边界的那一边，就归为哪一类。