# 生成模型与判别模型

生成模型与判别模型是针对于监督学习的划分，但同时无监督学习领域也有生成模型的说法。

监督学习可以分为两类：1）生成模型（generative model）；2）判别模型（Discriminative model）。

机器学习方法可以分为生成方法（generative approach）和判别方法（discriminative approach），所学到的模型分别称为生成式模型（generative model）和判别式模型（discriminative model）。

## 两者的比较

### 例子一

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。这就是GM----生成模型。

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。这就是DM----判别模型。

### 例子二

对于跟踪算法而言

1）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。如：argmax。

2）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。

### 例子三

1）判别式模型，这种模型的形式主要是根据原始图像推测图像具备的一些性质，例如根据数字图像推测数字的名称，根据自然场景图像推测物体的边界；

2）生成模型恰恰相反，通常给出的输入是图像具备的性质，而输出是性质对应的图像。这种生成模型相当于构建了图像的分布，因此利用这类模型，我们可以完成图像自动生成（采样）、图像信息补全等工作。

### 区别

1）生成方法的特点

生成方法可以还原出联合概率分布P(X,Y)，而判别方法则不能；

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。

2）判别方法的特点

判别方法直接学习的是条件概率P(Y|X)或决策函数f(X)，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f(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。

### 另一种定义

生成模型是已知样本的标签值y，对样本的特征向量x的条件概率进行建模，即对条件概率p(x|y)建模，它研究的是每种样本服从何种概率分布。

判别模型则刚好相反，已知样本的特征向量x，对样本的标签值y的概率进行建模，即对条件概率p(y|x)建模，这种一般用于分量，即给定样本x，计算它属于每个类的概率。

## 1. 生成模型

生成方法通过观测数据学习样本与标签的联合概率分布P(X, Y)，训练好的模型能够生成符合样本分布的新数据，它可以用于有监督学习和无监督学习。生成式方法，即概率方法，它完全可以从贝叶斯的角度进行理解。

总的来说，生成模型对联合概率p(x, y)建模，然后可以得到类后验概率p(y|x)。

本质上，生成模型是对联合概率分布p(x,y)=p(x|y)p(y)进行学习的。实际上，生成模型在学习过程中可以得到联合分布，但其最终学习的目标可以不是联合分布，可以与判别模型一样学习条件概率分布（生成模型一般也是通过最大化后验概率的形式进行建模优化）。联合分布虽然能提供更多的信息，但也需要更多的样本和更多的计算。因此，对于样本数量较少的问题，优先考虑判别模型。

所列举的Generative model也可以用disriminative方法来训练，比如GMM或HMM，训练的方法有EBW(Extended Baum Welch),或最近Fei Sha提出的Large Margin方法。

传统的生成模型包括：朴素贝叶斯，隐马尔可夫等，HMM等。

例如：生成模型的典型代表是贝叶斯分类器，它对类条件概率p(x|y)建模，而p(x|y)p(y)就是联合概率p(x, y)。通过贝叶斯公式，根据联合概率又可以得到类后验概率。

在深度学习之前已经有很多生成模型，但苦于生成模型难以描述难以建模，科研人员遇到了很多挑战，而深度学习的出现帮助他们解决了不少问题。深度生成模型有很多种形式，比如变分自编码器（VAE）、生成对抗网络（GAN）、自回归模型（PixelCNN）和可逆隐变量模型（Glow）。后两类模型尤其吸引研究者的注意力，因为它们提供了对边际似然度（marginal likelihood）的精确计算，且无需近似推断技术。

分类中的生成式方法和判别式方法已经为大家所熟知，其实降维中也有生成式方法和非生成式方法。

## 2. 判别模型

判别模型其实可以分为两种：一种是从概率角度出发（对应于生成模型），另一种是和概率完全没有关系的算法。

### 2.1 从概率角度出发的判别模型

判别方法由数据直接学习决策函数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，判别方法关心的是对给定的输入X，应该预测什么样的输出Y。

最典型的就是logistic回归和softmax回归，它们直接对p(y|x)建模，而不对p(x, y)建模。

### 2.2 和概率无关的判别模型

分类器根本就不建立概率模型，而是直接得到分类结果，这种是非概率模型，也称为判别模型。它直接根据样本x预测出类别y，这类模型没有使用概率的观点进行建模，而是用几何或者分析（函数）的手段建模，或者说从几何/函数角度出发，如找出分类超平面或者曲面，直接得到映射函数。

这类模型的典型代表是决策树，支持向量机（SVM），决策树/随机森林，kNN算法（K近邻法），AdaBoost算法，xgboost，标准的人工神经网络（包括全连接神经网络，卷积神经网络，循环神经网络等）、最大熵模型、CRF等等。

如果神经网络的最后一层是softmax变换，即softmax回归，则可以认为有概率的思想在其中；如果没有使用，则是概率无关。

支持向量机自始至终没有假设样本向量x服从何种分布，也没有估计类后验概率p(y|x)。这可以看成是一种几何划分的思想，把空间划分成多个部分。

决策树的预测函数时分段常数函数，直接实现从向量x到类别标签y的映射，没有计算任何概率值。其他的算法如随机森林，kNN，也是如此。

## 参考资料

https://blog.csdn.net/sigai\_csdn/article/details/83000946

https://www.cnblogs.com/szxspark/p/8426850.html

https://blog.csdn.net/dulingtingzi/article/details/80267726

http://blog.sciencenet.cn/home.php?mod=space&uid=248173&do=blog&id=227964

http://blog.163.com/huai\_jing@126/blog/static/1718619832011227757554/