

机器学习基础 | 大话生成模型与判别模型

原创 石头 机器学习算法那些事 2019-12-19

监督学习包含了两类构建机器学习模型的方法——判别模型和生成模型，我们以一篇生动的故事来引出这两类模型的概念。

故事：

有一个父亲有两个孩子，孩子A有能够深入学习一切事物的特殊才能，孩子B有能够学习他所看到事物之间区别的能力。

一天，父亲带两个孩子去动物园，这个动物园很小且只有两类动物——狮子和大象。他们离开动物园后，父亲指向一只动物并问他们：这只动物是狮子还是大象？

孩子A根据自己在动物园里看到狮子和大象的特征，马上在一张纸上画出了狮子和大象，然后他将这两幅图像与站在他面前的动物进行了比较，并根据与动物最接近的图像匹配作出了回答：“这只动物是狮子”。

孩子B根据学习到的可以区分大象和狮子的特征，作出了回答：“这只动物是狮子”。

在这里，他们的任务都是对眼前的动物进行分类，但是学习和寻找答案的方式是完全不同的，在机器学习中，**我们通常把孩子A称为生成模型，孩子B称为判别模型**。

生成模型学习联合概率分布 $p(x,y)$ ，然后根据贝叶斯理论预测条件概率 $p(y|x)$ ，推导公式为：

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)} = \frac{p(x,y)}{\sum_y p(x,y)} \quad (1)$$

判别模型直接学习条件概率分布 $P(y|x)$ 或学习决策函数 $y=f(x)$ 预测输出。

由 (1) 式可知生成模型可以得到判别模型，但判别模型得不到生成模型。

如何用生成模型和判别模型进行分类？**我们知道生成模型和判别模型的分类方法都是基于条件概率分布 $p(y|x)$ 进行的**，对于K分类，我们选择条件概率最大时的类为预测类，即：

$$y_{pred} = \arg \max_{C_k} [p(C_k | x)] \quad k=1,2,\dots,K \quad (2)$$

再次回顾 (1) 式，我们发现 $p(x)$ 是一个常数，因此 $p(y|x)$ 与 $p(x,y)$ 是成正比的关系，即：

$$p(y|x) \propto p(x,y) \quad (3)$$

下面举一个简单的例子来解释生成模型和判别模型的构建过程以及如何用该模型去进行分类：

假设我们进行二分类，训练数据的特征 x 有两个取值——1和2，输出 y 有两类——0或1，有如下的训练样本
 $(1,0)$ 、 $(1,0)$ 、 $(1,1)$ 、 $(2,1)$

由训练样本学习到的生成模型的联合概率分布如下：

$x \backslash y$	0	1
1	1/2	1/4
2	0	1/4

而学习到的判别模型的条件概率分布如下：

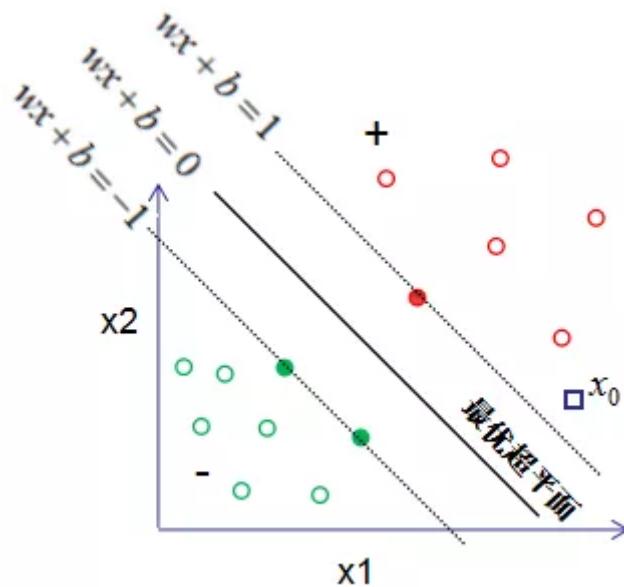
$x \backslash y$	0	1
1	2/3	1/3
2	0	1

若测试数据的输入特征 x 为1，根据(2)式(3)式，判别模型和生成模型的预测类别都是0。

下面来盘点下常见机器学习方法的模型分类：

1. 判别式分类模型：支持向量机

支持向量机的算法思想是寻找最大化不同类的函数间隔的超平面，如下图根据训练数据求解二分类的最优超平面：



判别函数 $f(x)$ 为：

$$f(x) = wx + b$$

给定一个测试数据 x_0 （上图方框表示），代入判别函数 $f(x)$ ：

$$f(x_0) = w x_0 + b > 1$$

因此测试数据标记为正样本。

2.生成式分类模型：朴素贝叶斯分类器

假设M维输入特征向量 $x = (x_1, x_2, \dots, x_M)$ ，目标向量 y 。给定训练数据，我们构建联合概率分布 $p(x, y)$ 。

$$p(x, y) = p(y) * p(x | y)$$

由条件独立性假设可得：

$$p(x, y) = p(y) * \prod_{m=1}^M p(x_m | y)$$

对上式积分可得：

$$p(x) = \sum_y p(x, y)$$

由贝叶斯理论可得后验概率：

$$p(y | x) = \frac{p(x, y)}{p(x)}$$

因此对于测试数据，分类结果为后验概率最大的输出 y 。

3.判别式分类模型：逻辑斯蒂回归

对于特征向量 x 的二分类（C1类和C2类）任务，逻辑斯蒂回归的核心是预测输入特征 x 的所属类的后验概率 $p(C_1 | x)$ ，可写为：

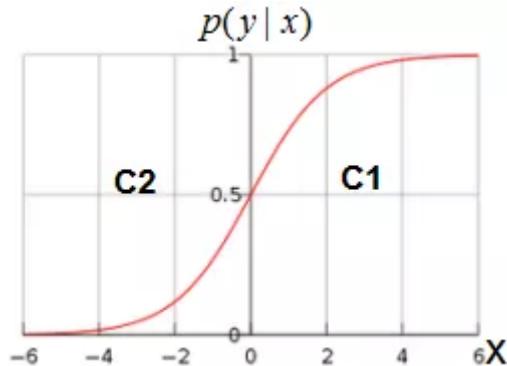
$$p(C_1 | x) = f(x) = \sigma(w^T x)$$

其中：

$$\sigma(a) = \frac{1}{1 + \exp(-a)} \quad (3.1)$$

$p(C_1 | x) \geq \frac{1}{2}$
若，则判断该输入特征属于C1类，否则输入C2类；

如下图：



逻辑斯蒂回归模型是直接预测条件概率 $p(y|x)$ 的分布，因此逻辑斯蒂回归是判别式分类模型。

4. 逻辑斯蒂回归模型与贝叶斯分类器的比较

上一节介绍逻辑斯蒂回归算法直接预测条件概率 $p(y|x)$ 的分布，给出分类结果。

下面看看贝叶斯是如何计算条件概率：

$$\begin{aligned}
 p(C_1 | x) &= \frac{p(x, C_1)}{p(x, C_1) + p(x, C_2)} \\
 &= \frac{p(x | C_1)p(C_1)}{p(x | C_1)p(C_1) + p(x | C_2)p(C_2)} \\
 &= \frac{1}{1 + \exp(-a)} \\
 &= \sigma(a)
 \end{aligned}$$

其中：

$$a = \ln \frac{p(x | C_1)p(C_1)}{p(x | C_2)p(C_2)} \quad (4.1)$$

比较式 (3.1) 和 (4.1) 式，可知两种模型参数 a 的求解方法不同：

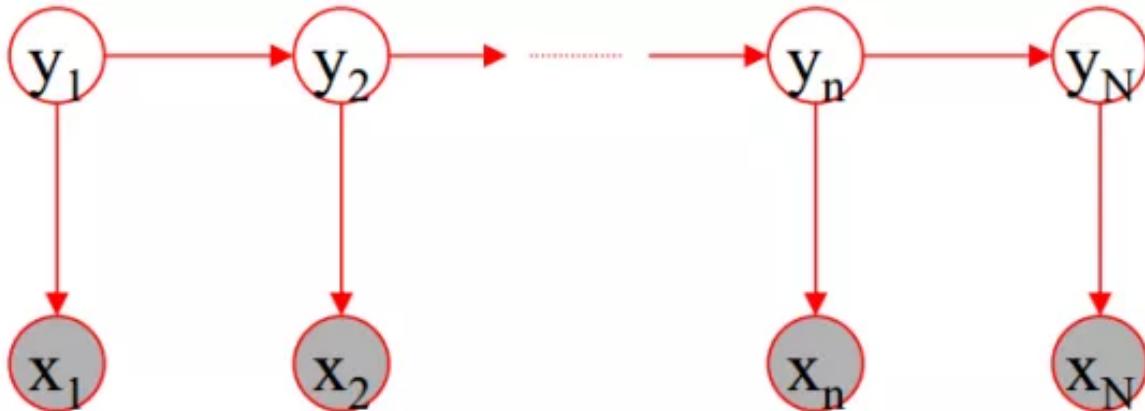
逻辑斯蒂回归模型： a 是输入特征 x 的线性方程，即：

$$a = w^T x$$

贝叶斯模型：我们首先估计模型的类条件概率，然后由 (4.1) 式计算得到参数 a ；

5. 生成模型：隐马尔科夫模型 (HMM)

给定观测数据序列 $X = \{x_1, x_2, \dots, x_N\}$, 如何预测生成观测数据的状态序列 $Y = \{y_1, y_2, \dots, y_N\}$, 观测数据序列和状态序列的关系如下图:



HMM通过构建联合概率分布 $P(X,Y)$ 来预测状态序列 Y , 联合概率分布为:

$$p(X, Y) = \prod_{n=1}^N p(y_n | y_{n-1}) p(x_n | y_n) \quad (5.1)$$

然后通过EM算法求解模型参数, 预测状态序列。

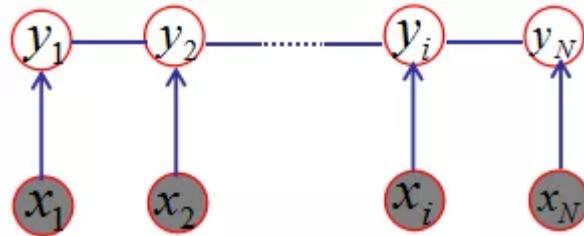
(5.1) 式成立时的两个假设条件:

- 过去与未来的状态相互独立, 即 t 时刻的状态只依赖于其前一时刻的状态, 与其他时刻的状态无关;
- 各状态的观测数据是相互独立的;

隐马尔可夫模型是通过联合概率分布 $P(X,Y)$ 来预测状态序列, 因此隐马尔可夫模型是生成模型。

6. 判别模型: 条件随机场 (CRF)

为了便于理解, 本节只讨论线性链条件随机场, 且假设观测序列 $X = \{x_1, x_2, \dots, x_N\}$ 和状态序列 $Y = \{y_1, y_2, \dots, y_N\}$ 有相同的结构, 如下图:



CRF是通过构建条件概率分布 $p(Y|X)$ 预测状态序列, 条件概率分布为:

$$p(Y | X) = \frac{1}{Z(X)} \exp \left\{ \sum_{m=1}^M \lambda_m f_m(y_n, y_{n-1}, x_n) \right\} \quad (6.1)$$

其中 $f_m(y_n, y_{n-1}, x_n)$ 是全局特征方程， λ_m 是对应的权值。

$Z(X)$ 是标准化方程，即：

$$Z(X) = \sum_y \exp \left\{ \sum_{m=1}^M \lambda_m f_m(y_n, y_{n-1}, x_n) \right\}$$

然后通过极大化训练数据的对数似然函数求解模型参数，预测状态序列。

7. 生成模型与判别模型的特点比较

生成模型是学习联合概率分布 $\tilde{p}(x, y)$ ，然后利用贝叶斯理论去预测条件概率分布 $p(y|x)$ 。生成模型明确的反映了数据集的分布情况，根据数据集的分布情况去预测测试数据的分类情况，当数据量较少时，学习到的联合概率分布 $\tilde{p}(x, y)$ 与真实的联合概率分布 $p(x, y)$ 差异较大，因此生成模型适合在数据量很大的情况，当模型需要考虑隐变量时，仍可用生成模型方法，此判别方法不能用。

判别模型是直接学习判别函数 $f(x)$ 或条件概率分布 $p(y|x)$ ，它是根据训练数据不同类别之间的特征差异来学习决策边界，如支持向量机，逻辑斯蒂回归等模型。因此在数据量不多的情况下判别模型的分类结果要好于生成模型，但是判别模型不能反应数据集的真实分布情况。

8. 总结

生成模型和判别模型一般处理监督学习的问题，这两类模型都是通过预测条件概率分布 $p(y|x)$ 进行分类（判别模型也可通过决策函数分类），只是方式不同：

生成模型思想：

- 1) 学习联合概率分布 $p(x, y)$ ；
- 2) 贝叶斯理论计算条件概率 $p(y|x)$ ；

判别模型思想：

- 1) 直接通过训练数据估计条件概率分布 $P(y|x)$ 或决策函数 $f(x)$ ；

生成模型的分类器：

- 朴素贝叶斯
- 贝叶斯网络
- 马尔科夫随机场
- 隐马尔可夫模型

判别模型的分类器：

逻辑斯蒂回归
支持向量机
最近邻算法
条件随机场

推荐阅读

520 页机器学习笔记！图文并茂可能更适合你，文末附下载方法

李航老师《统计学习方法》(第2版) 课件分享，文末附下载

Github | 吴恩达新书《Machine Learning Yearning》完整中文版开源

经典好书 | 141页的《Deep Learning with PyTorch》开源书籍

欢迎扫码关注，一起学习进步：

