

朴素贝叶斯分类器和一般的贝叶斯分类器有什么区别？

如题。或者说朴素贝叶斯的朴素在什么地方，和一般贝叶斯有什么区别呢？谢谢。

忆臻 也关注了该问题

添加评论 分享 邀请回答 举报

关注问题 写回答

16 个回答默认排序



茉茉
有色

12 人赞同了该回答

只要是基於貝葉斯理論的分類器就可以叫貝葉斯分類器，朴素貝葉斯分類器的叫法是因為它是優化過的一種運算性能高的算法（wikipedia 上的英語頁只給了朴素貝葉斯分類其的定義）。

貝葉斯網絡是一種基於貝葉斯理論以 DAG 形式描述全局概率分佈的一種統計方法，不屬於分類器的一種，主要用於貝葉斯推斷。

发布于 2014-05-11

12 7 条评论 分享 收藏 感谢



禾惠

忆臻 等 18 人赞同了该回答

从分类的角度说说自己的理解~

最大的不同是朴素贝叶斯假设特征之间是独立，互不影响。

拿分类问题举例，对于样本x(假设是一个n维的向量)，我们要求 $p(y|x)$ ，来估计所属的类别y。

根据贝叶斯公式，可以得到：

$$p(y|x) = p(x|y)p(y)/p(x) \dots\dots$$

右边的这个 $p(x|y)$ 是我们希望通过训练样本求得的。

如果是普通的贝叶斯，那么这个 $p(x|y)$ 的空间会很大，与n的大小乘指数级增长(因为有 2^n 种可能嘛)；

如果是朴素贝叶斯，既然条件独立，那么就可以把我们要求解的这个式子写成下面的样子：

$$p(x|y) = p(x_1, x_2, \dots | y) = p(x_1|y) * p(x_2|y) * \dots$$

这样一来，就可以把hypothesis的空间减少到与n的大小是线性关系

欢迎指正！

编辑于 2015-12-16

18 2 条评论 分享 收藏 感谢



阴影线
web dev

19 人赞同了该回答

本来我还想回（zhuang）答（ge）下（bi）的，可一看，上面各路大神已经回答得差不多了，真（hao）佩（xin）服（sai）。。。

我还是厚着脸皮说几句吧，朴素贝叶斯（naive Bayes）法是基于贝叶斯定理和特征条件独立假设的分类方法，对于给定的训练数据集，首先基于特征条件独立假设学习输入/输出的联合分布概率；然后基于此模型，对给定的输入x，再利用贝叶斯定理求出其后验概率最大的输出y。

1，它是一种典型的生成学习方法，其生成方法是由训练数据学习联合概率分布P(X,Y)，具体来说就是利用训练数据学习P(Y)和P(X|Y)的估计： $P(X,Y) = P(Y)P(X|Y)$ ，其概率估计方法是极大似然估计或者贝叶斯估计都行~~~

2，另一个要注意的点是朴素贝叶斯的基本假设是条件独立性。

相关问题

用于数据挖掘的分类算法有哪些，各有何优劣？8 个回答

怎样用非数学语言讲解贝叶斯定理（Bayes' theorem）？44 个回答

数据挖掘、人工智能、模式识别等学科的公共数学基础有哪些？（比如贝叶斯网络）17 个回答

豆瓣九点是如何做文章自动分类的？6 个回答

有哪些关于智能算法的好书籍？5 个回答

相关 Live 推荐

《线性代数》中最重要的矩阵
Heshawn
★★★★★ 76 人参与

数据达人养成计划「实战篇」
Han Hsiao
★★★★★ 103 人参与

线性代数入门：几个基本概念
胡鞍钢
★★★★★ 513 人参与

刘看山 · 知乎指南 · 知乎协议 · 应用 · 工作

侵权举报 · 网上有害信息举报专区

违法和不良信息举报：010-82716601

儿童色情信息举报专区

联系我们 © 2017 知乎

$$P(X = x | Y = c_k) = P(X^{(1)} = x^{(1)}, \dots, X^{(n)} = x^{(n)} | Y = c_k) \\ = \prod_{j=1}^n P(X^{(j)} = x^{(j)} | Y = c_k)$$

(因为网上没有找到这个公式, 所以我自己打开了word~自从数学建模后就没编辑过这么复杂的公式, 2333~)

这会使模型包含的条件概率的数量减少, 因而朴素贝叶斯法高效, 易于实现, 但也有缺点, 就是其分类的性能不一定很高。

3, 朴素贝叶斯法利用贝叶斯定理与学到的联合概率模型进行分类预测,

$$P(Y | X) = \frac{P(X, Y)}{P(X)} = \frac{P(Y)P(X | Y)}{\sum_Y P(Y)P(X | Y)}$$

将输入x分到后验概率最大的类y.

$$Y = \arg \max P(Y = c_k) \prod_{j=1}^n P(X_j = x^j | Y = c_k)$$

后验概率最大等价于0-1损失函数时的期望风险最小化。。。

参考文献:《统计学习基础》范明 电子工业版 2004

《统计学习方法》李航 清华大学出版社 2014

如果回答得不对请告诉我, 反正我也不会改, 我只是个前端~~~

发布于 2015-09-04

▲ 19 ▼

3 条评论

分享

★ 收藏

♥ 感谢

...

收起 ^



忆臻

PHD Candidate

32 人赞同了该回答

我根据自己理解, 给出朴素贝叶斯分类算法的核心, 在里面会有通俗详细的解释, 希望对你有帮助~

贝叶斯分类是一类分类算法的总称, 这类算法均以贝叶斯定理为基础, 故统称为贝叶斯分类。而朴素贝叶斯分类是贝叶斯分类中最简单, 也是常见的一种分类方法。

分类问题综述

对于分类问题, 其实谁都不会陌生, 日常生活中我们每天都进行着分类过程。例如, 当你看到一个人, 你的脑子下意识判断他是学生还是社会上的人; 你可能经常会走在路上对身旁的朋友说“这个人一看就很有钱。”之类的话, 其实这就是一种分类操作。

既然是贝叶斯分类算法, 那么分类的数学描述又是什么呢?

从数学角度来说, 分类问题可做如下定义: 已知集合 $C = y_1, y_2, \dots, y_n$ 和

$I = x_1, x_2, x_3, \dots, x_n$, 确定映射规则 $y = f(x)$, 使得任意 $x_i \in I$ 有且仅有一个 $y_i \in C$, 使得 $y_i \in f(x_i)$ 成立。

其中C叫做类别集合, 其中每一个元素是一个类别, 而叫做项集合(特征集合), 其中每一个元素是一个待分类项, f叫做分类器。分类算法的任务就是构造分类器f。

分类算法的内容是要求给定特征, 让我们得出类别, 这也是所有分类问题的关键。那么如何由指定特征, 得到我们最终的类别, 也是我们下面要讲的, 每一个不同的分类算法, 对应着不同的核心思想。

在回答中, 我会用一个具体实例, 对朴素贝叶斯算法几乎所有的重要知识点进行讲解。

朴素贝叶斯分类

那么既然是朴素贝叶斯分类算法, 它的核心算法又是什么呢?

是下面这个贝叶斯公式:

换个表达形式就会明朗很多，如下：

我们最终求的 $p(\text{类别}|\text{特征})$ 即可！就相当于完成了我们的任务。

例题分析

下面我先给出例子问题。

给定数据如下：

现在给我们的问题是，如果一对男女朋友，男生想女生求婚，男生的四个特点分别是不帅，性格不好，身高矮，不上进，请你判断一下女生是嫁还是不嫁？

这是一个典型的分类问题，转为数学问题就是比较 $p(\text{嫁}|\text{不帅、性格不好、身高矮、不上进})$ 与 $p(\text{不嫁}|\text{不帅、性格不好、身高矮、不上进})$ 的概率，谁的概率大，我就能给出嫁或者不嫁的答案！

这里我们联系到朴素贝叶斯公式：

我们需要求 $p(\text{嫁}|\text{不帅、性格不好、身高矮、不上进})$ ，这是我们不知道的，但是通过朴素贝叶斯公式可以转化为好求的三个量， $p(\text{不帅、性格不好、身高矮、不上进}|\text{嫁})$ 、 $p(\text{不帅、性格不好、身高矮、不上进})$ 、 $p(\text{嫁})$ （至于为什么能求，后面会讲，那么就太好了，将待求的量转化为其它可求的值，这就相当于解决了我们的问题！）

朴素贝叶斯算法的朴素一词解释

那么这三个量是如何求得？

是根据已知训练数据统计得来，下面详细给出该例子的求解过程。

回忆一下我们要求的公式如下：

那么我只要求得 $p(\text{不帅、性格不好、身高矮、不上进}|\text{嫁})$ 、 $p(\text{不帅、性格不好、身高矮、不上进})$ 、 $p(\text{嫁})$ 即可，好的，下面我分别求出这几个概率，最后一比，就得到最终结果。

$p(\text{不帅、性格不好、身高矮、不上进}|\text{嫁}) = p(\text{不帅}|\text{嫁}) * p(\text{性格不好}|\text{嫁}) * p(\text{身高矮}|\text{嫁}) * p(\text{不上进}|\text{嫁})$ ，那么我就要分别统计后面几个概率，也就得到了左边的概率！

等等，为什么这个成立呢？学过概率论的同学可能有感觉了，**这个等式成立的条件需要特征之间相互独立吧！**

对的！这也就是为什么朴素贝叶斯分类有朴素一词的来源，朴素贝叶斯算法是假设各个特征之间相互独立，那么这个等式就成立了！

但是为什么需要假设特征之间相互独立呢？

1、我们这么想，假如没有这个假设，那么我们对右边这些概率的估计其实是不可做的，这么说，我们这个例子有4个特征，其中帅包括{帅，不帅}，性格包括{不好，好，爆好}，身高包括{高，矮，中}，上进包括{不上进，上进}，那么四个特征的联合概率分布总共是4维空间，总个数为 $2*3*3*2=36$ 个。

24个，计算机扫描统计还可以，但是现实生活中，往往有非常多的特征，每一个特征的取值也是非常之多，那么通过统计来估计后面概率的值，变得几乎不可做，这也是为什么需要假设特征之间独立的原因。

2、假如我们没有假设特征之间相互独立，那么我们统计的时候，就需要在整个特征空间中去找，比如统计 $p(\text{不帅、性格不好、身高矮、不上进}|\text{嫁})$ ，

我们就需要在嫁的条件下，去找四种特征全满足分别是不帅，性格不好，身高矮，不上进的人的个数，这样的话，由于数据的稀疏性，很容易统计到0的情况。这样是不合适的。

根据上面两个原因，朴素贝叶斯法对条件概率分布做了条件独立性的假设，由于这是一个较强的假设，朴素贝叶斯也由此得名！这一假设使得朴素贝叶斯法变得简单，但有时会牺牲一定的分类准确率。

好的，上面我解释了为什么可以拆成分开连乘形式。那么下面我们就开始求解！

到这里为止，给出了问题中朴素贝叶斯分类和一般的贝叶斯分类的区别。

其它朴素贝叶斯分类内容见：[带你理解朴素贝叶斯分类算法 - 知乎专栏](#)

编辑于 2017-05-01

▲ 32

● 2 条评论

➦ 分享

★ 收藏

♥ 感谢

...

收起 ^



范文

推荐系统/数据挖掘/文本挖掘/情感倾向性分析/自然语言处理

13 人赞同了该回答

条件独立性和位置无关性

发布于 2013-04-15

▲ 13

● 1 条评论

➦ 分享

★ 收藏

♥ 感谢

...



肖智博

新加坡南洋理工大学 电气与电子工程博士后

12 人赞同了该回答

“朴素”是说NB要满足**条件独立性**假设。而且，好像没听说过一般的贝叶斯分类器这个说法，我猜你指的是贝叶斯网络吧。个人理解，NB和BN最明显的区别就在条件独立性了。BN有一大套学习结构的方法，NB相比较而言就算是“朴素”了吧。

发布于 2012-03-24

▲ 12

▼

💬 1 条评论

➦ 分享

★ 收藏

♥ 感谢

⋮



池吉

10 人赞同了该回答

四个字：条件独立。

朴素是英文naive的翻译，之所以说NB模型naive，是因为它假设了特征（ $X_{i:1-j}$ ）在特定class label（ $y=c:1-C$ ）下的条件独立。对于实际应用，这显然是个很理想的状态。但有趣的是，尽管这种建模方式不贴近现实情况，但是在很多现实应用中，这种“naive”模型却能很好的解决问题，譬如说垃圾邮件识别问题。

当然了，朴素也有其好处。避免模型过于复杂，降低模型维数，可以有效防止过拟合（overfitting），并且节约模型的计算成本。

用已故的统计学家George E. P. Box的话来说，就是：All models are wrong, but some are useful.

编辑于 2016-07-03

▲ 10

▼

💬 添加评论

➦ 分享

★ 收藏

♥ 感谢

⋮



DanielLaah

机器学习, 推荐系统

8 人赞同了该回答

李航，统计学习方法。

发布于 2016-09-11

▲ 8

▼

💬 添加评论

➦ 分享

★ 收藏

♥ 感谢

⋮



puyangsky

我决定要好好玩了

3 人赞同了该回答

[朴素贝叶斯分类器的应用 - 阮一峰的网络日志](#) 阮一峰的博客还是非常容易看懂的

发布于 2016-11-25

▲ 3

▼


💬 添加评论

➦ 分享

★ 收藏

♥ 感谢

⋮



what

一只来自电子与通信工程的研究僧

2 人赞同了该回答

用一栗子来说下：

假设单词bacon出现在unhealthy后面与出现在delicious后面的概率相同。当然，这种假设并不正确，bacon常常出现在delicious附近，而很少出现在unhealthy附近，这个假设正是朴素贝叶斯分类器中朴素的含义。

即假设一个特征或者单词出现的可能性与他和其他单词相邻没有关系。

发布于 2015-11-18

▲ 2 ▼

● 添加评论

➦ 分享

★ 收藏

♥ 感谢

⋮



王子铭

1 人赞同了该回答

尽管实际上独立假设常常是不准确的，但朴素贝叶斯分类器的若干特性让其在实践中能够取得令人惊奇的效果。特别地，各类条件特征之间的解耦意味着每个特征的分布都可以独立地被当做一维分布来估计。这样减轻了由于维度灾难带来的阻碍,当样本的特征个数增加时就不需要使样本规模呈指数增长。然而朴素贝叶斯在大多数情况下不能对类概率做出非常准确的估计，但在许多应用中这一点并不要求。

编辑于 2017-02-12

▲ 1 ▼

● 添加评论

➦ 分享

★ 收藏

♥ 感谢

⋮



梁雨
机器不学习

1 人赞同了该回答

朴素，是因为做了一个假设：用于分类的特征是条件独立的。至于为什么要做这个假设，个人理解是计算上的方便，减小模型复杂度，如果不做这个假设，需要学习的条件概率是指数级数量的参数，如果假设条件独立，由于 $P(AB) = P(A) * P(B)$,参数数量会大大减小。

发布于 2016-11-24

▲ 1 ▼


● 添加评论

➦ 分享

★ 收藏

♥ 感谢

⋮



张西涛
架构师 移动互联网

条件独立性

发布于 2016-11-17

▲ 0 ▼


● 添加评论

➦ 分享

★ 收藏

♥ 感谢

⋮



JokerKing

基于条件独立性的假设。

发布于 2016-02-21

▲ 0 ▼

● 添加评论

➦ 分享

★ 收藏

♥ 感谢

⋮



Patrick Leung
ML/DM/NLP爱好者+Fitter

1 人赞同了该回答

1. 满足贝耶斯定理
2. 特征之前权重相等
3. 特征之前条件独立，条件独立，条件独立（最重要的压轴说三遍）

发布于 2015-09-05

▲ 1

▼

添加评论

分享

收藏

感谢

...



赛先生

人生如飞禽走兽一般 一千零一只猴子

计算方法不一样 根据训练效果来决定使用哪种分类法

发布于 2014-05-11

▲ 0

▼

添加评论


分享

收藏

感谢

...

1 个回答被折叠（为什么？）



Mr.Better

学生 [编辑话题经验](#)

使用匿名身份回答

B I H “ </> ☰ ☷ 🔗 🖼️ 📺 ∑ ☰ ✂️

↗️ 全屏模式

写回答...