# 机器学习——朴素贝叶斯 🙉

来自【机器学习面试题汇总与解析(蒋豆芽面试题总结)】 51 浏览 0 回复 2021-06-25



蒋豆芽 👜



# 机器学习面试题汇总与解析——朴素贝叶斯

- 1. **朴素贝叶斯介绍一下** $\diamondsuit$  $\diamondsuit$  $\diamondsuit$  $\diamondsuit$
- 2. **朴素贝叶斯优缺点**  $\diamondsuit$   $\diamondsuit$   $\diamondsuit$   $\diamondsuit$
- 3. **贝叶斯公式** ☆ ☆ ☆ ☆
- 4. 朴素贝叶斯中的"朴素"怎么理解?  $\Diamond \Diamond \Diamond \Diamond \Diamond$
- 5. **什么是拉普拉斯平滑法?** ☆ ☆ ☆ ☆
- 6. 朴素贝叶斯中有没有超参数可以调?  $\Diamond \Diamond \Diamond \Diamond \Diamond$
- 7. **你知道朴素贝叶斯有哪些应用吗?** ☆ ☆ ☆ ☆ ☆
- 8. **朴素贝叶斯对异常值敏不敏感?**  $\diamondsuit \diamondsuit \diamondsuit \diamondsuit \diamondsuit$
- 9. **频率学派与贝叶斯学派的差别**  $\diamondsuit$   $\diamondsuit$   $\diamondsuit$
- 10. 概率与期望的公式  $\Diamond$   $\Diamond$   $\Diamond$
- 11. 先验概率与后验概率 ☆ ☆ ☆ ☆

\_\_\_\_\_

- 本专栏适合于Python已经入门的学生或人士,有一定的编程基础。
- 本专栏适合于**算法工程师、机器学习、图像处理求职**的学生或人士。
- 本专栏针对面试题答案进行了**优化,尽量做到好记、言简意赅。这才是一份面试题总结的正确打开** 方式。这样才方便背诵
- 如专栏内容有错漏,欢迎在评论区指出或私聊我更改,一起学习,共同讲步。
- 相信大家都有着高尚的灵魂,请尊重我的知识产权,未经允许严禁各类机构和个人转载、传阅本专栏的内容。

关于**机器学习算法**书籍,我强烈推荐一本**《百面机器学习算法工程师带你面试》**,这个就很类似面经,还有讲解,写得比较好。私聊我进群。

关于**深度学习算法**书籍,我强烈推荐一本**《解析神经网络——深度学习实践手册》**,简称CNN book,通俗易懂。私聊我进群。

B站机器学习视频: https://space.bilibili.com/10781175/channel/detail?cid=133301

朴素贝叶斯: https://zhuanlan.zhihu.com/p/26262151

读者可以把参考文章看看

### 个人理解

**朴素贝叶斯**理论很美,常用于分类,属于生成式模型,由联合分布概率得出分类结果。

### 朴素贝叶斯的优缺点

#### 优点:

- 1. 算法逻辑简单,易于实现
- 2. 分类过程中时空开销小,分类准确度高,速度快。
- 3. 朴素贝叶斯模型发源于古典数学理论,有稳定的分类效率。
- 4. 对缺失数据不太敏感, 算法也比较简单, 常用于文本分类。
- 5. 对小规模的数据表现很好, 能处理多分类任务, 适合增量式训练

#### 缺点:

- 1. 理论上,朴素贝叶斯模型与其他分类方法相比具有最小的误差率。但是实际上并非总是如此,这是因为朴素贝叶斯模型假设**属性之间相互独立**,这个假设在实际应用中往往是不成立的,在属性个数比较多或者属性之间相关性较大时,**分类效果不好**。
- 2. 需要知道**先验概率**,且先验概率很多时候是基于假设或者已有的训练数据所得的,这在某些时候可能会因为假设先验概率的原因出现分类决策上的错误。
- 1. **朴素贝叶斯介绍一下** $\diamondsuit$  $\diamondsuit$  $\diamondsuit$  $\diamondsuit$

#### 参考回答

**朴素贝叶斯**是一个基于**特征条件独立假设和贝叶斯原理**的一种**分类算法。朴素贝叶斯**通过训练数据得到X与y的**联合分布**;之后对于要预测的X,根据**贝叶斯公式**,输出**后验概率**最大的y。

**朴素贝叶斯**是一种**生成式**学习算法,其生成方法通过学习X,Y的联合分布来实现的。假设各个特征在给定y的情况下是相互独立的。

#### 答案解析

无。

### :■ 蔣豆芽

### 参考回答

### 朴素贝叶斯的优缺点

### 优点:

- 1. 算法逻辑简单,易于实现
- 2. 分类过程中时空开销小, 分类准确度高, 速度快。
- 3. 朴素贝叶斯模型发源于古典数学理论,有稳定的分类效率。
- 4. 对缺失数据不太敏感, 算法也比较简单, 常用于文本分类。
- 5. 对小规模的数据表现很好,能处理多分类任务,适合增量式训练

#### 缺点:

- 1. 理论上,朴素贝叶斯模型与其他分类方法相比具有最小的误差率。但是实际上并非总是如此, 这是因为朴素贝叶斯模型假设**属性之间相互独立**,这个假设在实际应用中往往是不成立的,在 属性个数比较多或者属性之间相关性较大时,**分类效果不好**。
- 2. 需要知道**先验概率**,且先验概率很多时候是基于假设或者已有的训练数据所得的,这在某些时候可能会因为假设先验概率的原因出现分类决策上的错误。

#### 答案解析

无。

3. **贝叶斯公式** ☆ ☆ ☆ ☆

#### 参考回答

$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{\sum_{j=1}^n P(B_j)P(A|B_j)}$$
 牛宾@蒋豆芽

公式中,事件Bi的概率为P(Bi),事件Bi已发生条件下事件A的概率为P(A | Bi),事件A发生条件下事件Bi的概率为P(Bi | A)。

#### 答案解析

无。

4. **朴素贝叶斯中的"朴素"怎么理解?**  $\Diamond \Diamond \Diamond \Diamond \Diamond \Diamond$ 

#### 参考回答

### :■ 蔣豆芽

属性也会存在性,所以说朴素贝叶斯是一种很"朴素"的算法。

#### 答案解析

无。

5. **什么是拉普拉斯平滑法?**  $\diamondsuit$   $\diamondsuit$   $\diamondsuit$   $\diamondsuit$ 

#### 参考回答

**拉普拉斯平滑法**是**朴素贝叶斯**中处理**零概率**问题的一种修正方式。在进行分类的时候,可能会出现某个属性在训练集中没有与某个类同时出现过的情况,如果直接基于朴素贝叶斯分类器的表达式进行计算的话就会出现**零概率**现象。为了避免其他属性所携带的信息被训练集中未出现过的属性值"抹去",所以才使用**拉普拉斯估计器**进行修正。具体的方法是:在分子上加1,对于先验概率,在分母上加上训练集中可能的类别数;对于条件概率,则在分母上加上第1个属性可能的取值数

#### 答案解析

无。

6. 朴素贝叶斯中有没有超参数可以调? ☆ ☆ ☆ ☆ ☆

### 参考回答

**朴素贝叶斯**是没有超参数可以调的,所以它不需要调参,朴素贝叶斯是根据训练集进行分类,分类 出来的结果基本上就是确定了的,拉普拉斯估计器不是朴素贝叶斯中的参数,不能通过拉普拉斯估 计器来对朴素贝叶斯调参。

#### 答案解析

无。

7. **你知道朴素贝叶斯有哪些应用吗?**  $\diamondsuit \diamondsuit \diamondsuit \diamondsuit \diamondsuit$ 

#### 参考回答

**朴素贝叶斯**的应用最广的应该就是在**文档分**类、**垃圾文本过滤**(如垃圾邮件、垃圾信息等)、**情感分析**(微博、论坛上的积极、消极等情绪判别)这些方面,除此之外还有多分类实时预测、**推荐系统**(贝叶斯与协同过滤组合使用)、**拼写矫正**(当你输入一个错误单词时,可以通过文档库中出现的概率对你的输入进行矫正)等。

#### 答案解析

### :■ 蔣豆芽

8. 朴素贝叶斯对异常值敏不敏感? ☆ ☆ ☆ ☆ ☆

#### 参考回答

**朴素贝叶斯**对**异常值**不敏感。所以在进行数据处理时,我们可以不去除异常值,因为保留异常值可以保持朴素贝叶斯算法的整体精度,而去除异常值则可能在进行预测的过程中由于失去部分异常值导致模型的泛化能力下降。

### 答案解析

无。

9. 频率学派与贝叶斯学派的差别 ☆ ☆ ☆

#### 参考回答

**频率学派** - Frequentist - Maximum Likelihood Estimation (MLE,最大似然估计)

**贝叶斯学派** - Bayesian - Maximum A Posteriori (MAP,最大后验估计)

**频率学派**和**贝叶斯学派**对世界的认知有本质不同:**频率学派**认为世界是确定的,有一个本体,这个本体的真值是不变的,我们的目标就是要找到这个**真值或真值所在的范围**;而**贝叶斯学派**认为世界是不确定的,人们对世界先有一个预判,而后通过观测数据对这个预判做调整,我们的目标是要找到**最优的描述这个世界的概率分布**。

### 答案解析

无。

10. 概率与期望的公式 ☆ ☆ ☆ ☆

#### 参考回答

略

#### 答案解析

无。

11. 先验概率与后验概率 ☆ ☆ ☆ ☆

#### 参考回答

用"瓜熟蒂落"这个因果例子,从概率 (probability) 的角度说一下,

么瓜熟的概率是多少。

后验和先验的关系可以通过**贝叶斯公式**来求。也就是: P (瓜熟 | 已知蒂落) = P (瓜熟) × P (蒂落 | 瓜熟) / P (蒂落)

### 答案解析

无。

资源分享 python 机器学习 算法工程师 春秋招 面试题 软件开发 面经 举报





...

### 相关专栏



已订阅

0条评论 ○↑ 默认排序 ~



没有回复

请留下你的观点吧~

发布

# / 牛客博客, 记录你的成长