

Artificial Intelligence HW6

王嵘晟 PB1711614

13.15

$$P(\text{阳性}|\text{患病}) = 0.99, P(\neg\text{阳性}|\neg\text{患病}) = 0.99, P(\text{患病}) = 0.0001$$
$$\text{所以 } P(\text{患病}|\text{阳性}) = \frac{P(\text{阳性}|\text{患病})P(\text{患病})}{P(\text{阳性})} = \frac{P(\text{阳性}|\text{患病})P(\text{患病})}{P(\text{阳性}|\text{患病})P(\text{患病}) + P(\text{阳性}|\neg\text{患病})P(\neg\text{患病})} = \frac{0.99 \times 0.0001}{0.99 \times 0.0001 + 0.01 \times 0.9999} = 0.0098$$

因为“这种病很罕见”使得患病概率比较低，导致检测阳性时确诊患病的条件概率也比较低，所以是个好消息。实际患病概率为 **0.0098**

13.18

a.

$$\text{由已知条件: } P(\text{真币}) = \frac{n-1}{n}, P(\text{假币}) = \frac{1}{n}, P(\text{正面}|\text{假币}) = 1, P(\text{正面}|\text{真币}) = \frac{1}{2}$$
$$\text{所以所求为 } P(\text{假币}|\text{正面}) = \frac{P(\text{正面}|\text{假币})P(\text{假币})}{P(\text{正面}|\text{假币})P(\text{假币}) + P(\text{正面}|\text{真币})P(\text{真币})} = \frac{1 \times \frac{1}{n}}{1 \times \frac{1}{n} + \frac{1}{2} \times \frac{n-1}{n}} = \frac{2}{n+1}$$

b.

$$P(\text{假币}|K\text{次正面}) = \frac{P(K\text{次正面}|\text{假币})P(\text{假币})}{P(K\text{次正面}|\text{假币})P(\text{假币}) + P(K\text{次正面}|\text{真币})P(\text{真币})} = \frac{1 \times \frac{1}{n}}{1 \times \frac{1}{n} + (\frac{1}{2})^k \times \frac{n-1}{n}} = \frac{2^k}{n+2^k-1}$$

c

$$\text{发生错误的概率} = P(K\text{次正面}|\text{真币})P(\text{真币}) = \frac{n-1}{2^k n}$$

13.22

a.

模型包括先验概率 $P(\text{category})$ 和条件概率 $P(\text{content}_i|\text{category})$ 对于每个分类类别 c , $P(\text{category}=c)$ 是文档全被归类为 c 的概率, $P(\text{content}_i = \text{true}|\text{category} = c)$ 是被归类为 c 的文档包含内容 content_i 的概率

b.

$P(\text{content}_i|\text{category}), p(\text{category})$ 都可以根据已知条件算出来, 再由 Bayes 公式, 分类时将 $P(\text{content}_i|\text{category})p(\text{category})$ 最大的值分为一类, 然后剩余的文档重复此操作, 即可完成分类一个新文档。

c.

不合理, 自然语言组成的文档不是上下文无关的, 所以单词之间不具有独立性, 所以条件独立性假设不合理。