

第7次作业

13.15

这种病越罕见，导致结果呈阳性的原因越可能是没得病却误诊断

由贝叶斯公式，得

$$\text{患病概率 } P = \frac{0.99 \times 0.00001}{0.9999 \times 0.01 + 0.99 \times 0.0001} = 0.98\%$$

13.18

(a)

由贝叶斯公式，得

$$P = \frac{\frac{1}{n}}{\frac{n-1}{n} * \frac{1}{2} + \frac{1}{n}} = \frac{2}{n+1}$$

(b)

$$\begin{aligned} P(\text{fake} | \text{head}^k) &= \alpha P(\text{head}^k | \text{fake}) P(\text{fake}) \\ &= \alpha < 1, 0.5^k > < \frac{1}{n}, \frac{n-1}{n} > \\ &= < \frac{2^k}{2^k + n - 1}, \frac{n-1}{2^k + n - 1} > \end{aligned}$$

(c)

发生错误的概率：

$$P = \frac{n-1}{2^k * n}$$

13.21

(a)

不可以，因为不知道蓝色车在出租车中的占比

如果蓝色车占比很少，事实是肇事车是绿色车，而目击者误判的概率更大

(b)

$$\begin{aligned} P(\text{blue} | \text{lookblue}) &= \alpha P(\text{lookblue} | \text{blue}) P(\text{blue}) \\ &= \alpha < 0.75, 0.25 > < 0.1, 0.9 > \text{故更可能是绿色车} \\ &= < 0.25, 0.75 > \end{aligned}$$

13.22

(a)

先验概率为 $P(category)$, 条件概率为 $P(word|category)$, 对于每个 $word_i$ 和种类 c , 都存在对应的概率 $P(u$

(b)

即给定所有 $word_i = true \text{ or } false$, 对每个 $category$, 计算 $P(category = c|e)$, 其中 e 为观测到的所有 $word_i$
这时, 我们可以使用贝叶斯公式计算每个分类的概率

$$P(category = c|e) = \alpha P(category = c, e) = \alpha P(e|category = c)P(category = c)$$

最后取其中的最大值作为类型

(c)

不合理, 因为在语言中, 不同的单词之间是有关联的, 不可能是独立的, 可以使用Ngram模型解决这个问题

14.12

(a)

我认为ii更正确的表示了, 观测结果M与数量N和对焦失误F之间的因果关系

iii也正确, 但更复杂, 因为它颠倒了因果

(b)

ii把原因放在了结果前面, 它的参数更少, 所以是最好的

(c)

$$\mathbf{P}(M_1|N) == \mathbf{P}(M_1|N, F_1)\mathbf{P}(F_1) + \mathbf{P}(M_1|N, \neg F_1)\mathbf{P}(\neg F_1)$$

其中 $P(F_1) = f$, 观测少一颗星的概率为 e , 多一颗星概率为 e

由上可得

	N=1	N=2	N=3
$M_1 = 0$	$f + (1 - f)e$	f	f
$M_1 = 1$	$(1 - f)(1 - 2e)$	$(1 - f)e$	0
$M_1 = 2$	$(1 - f)e$	$(1 - f)(1 - 2e)$	$(1 - f)e$
$M_1 = 3$	0	$(1 - f)e$	$(1 - f)(1 - 2e)$
$M_1 = 4$	0	0	$(1 - f)e$

(c)

由 $M_1 = 1$, 得 N 的可能取值为2, $[4, +\infty)$

由 $M_2 = 3$, 得 N 的可能取值为4, $[6, +\infty)$

所以, N 的可能取值为2, 4, $[6, +\infty)$, $N \in \mathbf{Z}$

(d)

易知, 当 $N \geq 6$ 时, 随着 N 的增大, 为 N 的后验概率越来越小, 所以只讨论 $N = 2, 4, 6$ 三种情况

$$P(N = 2|M_1 = 1, M_2 = 3) = \frac{P(M_1=1, M_2=3|N=2)P(N=2)}{P(M_1=1, M_2=3)} = (1 - f)^2 e^2 * P(N = 2) / P(M_1 = 1, M_2 = 3)$$

$$P(N = 4|M_1 = 1, M_2 = 3) = \frac{P(M_1=1, M_2=3|N=4)P(N=4)}{P(M_1=1, M_2=3)} = (1 - f)e * f * P(N = 4)/P(M_1 = 1, M_2 = 3)$$

(如果考虑N=4, 发生F, 导致M1=0的情况, 则P会更小)

$$P(N = 6|M_1 = 1, M_2 = 3) = \frac{P(M_1=1, M_2=3|N=6)P(N=6)}{P(M_1=1, M_2=3)} = f^2 * P(N = 6)/P(M_1 = 1, M_2 = 3)$$

(如果考虑N=6, 发生F, $M_1 \neq 1, M_2 \neq 3$ 的情况, 则P会更小)

如果我们假设先验概率 $P(N = k)$ 的概率都大致相同, 由于 $f \ll e$

$P(N = 2|M_1 = 1, M_2 = 3)$ 最大

所以在先验概率 $P(N = k)$ 的概率都大致相同的情况下,

恒星数目N=2的概率最大

14.13

$$\begin{aligned} \mathbf{P}(N|M_1 = 2, M_2 = 2) &= \alpha \sum_{f_1 f_2} P(N, M_1 = 2, M_2 = 2, f_1, f_2) \\ &= \alpha \sum_{f_1 f_2} P(N)P(f_1)P(f_2)P(M_1 = 2|N, f_1)P(M_2 = 2|N, f_2) \end{aligned}$$

由于 $N \in 1, 2, 3$

所以 $f_1 = false, f_2 = false$,只用枚举这一种情况

所以

$$\begin{aligned} \mathbf{P}(N|M_1 = 2, M_2 = 2) &= \alpha \sum_{f_1 f_2} P(N, M_1 = 2, M_2 = 2, f_1, f_2) \\ &= \alpha \sum_{f_1 f_2} P(N)P(f_1)P(f_2)P(M_1 = 2|N, f_1)P(M_2 = 2|N, f_2) \\ &= \alpha' < P(N = 1) * e^2, P(N = 2) * (1 - 2e)^2, P(N = 3) * e^2 > \end{aligned}$$