HW7

PB21111686_赵卓

13.15

- a.
“这种病很罕见”是好消息的原因是，如果这种病的发病率远远小于测量错误的概率，那么阳性是测量错误的可能就很大了，当然是好消息
- b.
人群发病率为0.0001
因此未患病误测的概率为$\frac{0.9999\times0.01}{0.0001\times0.99+0.9999\times0.01}\approx99.02\%$
而真实患病的概率为$\frac{0.0001\times0.99}{0.0001\times0.99+0.9999\times0.01}\approx0.98\%$

13.18

- a.
正面朝上的概率为$\frac{n-1}{n}\times\frac{1}{2}+\frac{1}{n}=\frac{n+1}{2n}$
因此取出伪币的概率为$\frac{\frac{1}{n}}{\frac{n+1}{2n}}=\frac{2}{n+1}$
- b.
抛k次都是正面朝上的概率为$\frac{n-1}{n}\times(\frac{1}{2})^k+\frac{1}{n}=\frac{n+2^k-1}{n2^k}$
因此取出伪币的概率为$\frac{\frac{1}{n}}{\frac{n+2^k-1}{n2^k}}=\frac{2^k}{n+2^k-1}$
- c.
发生错误的概率为取出真币但是k次都正面朝上的概率$\frac{n-1}{n}\times(\frac{1}{2})^k=\frac{n-1}{n2^k}$

13.21

- a.
有可能，根据题给条件，目击者所说的颜色可信度为75%，也就是肇事车的颜色确实是目击者所说的可能性是75%，因此最可能是蓝色。
- b.
如果10辆中有9辆是绿色的，那么看成蓝色的概率为$\frac{9}{10}\times\frac{1}{4}+\frac{1}{10}\times\frac{3}{4}=\frac{3}{10}$
因此确实是蓝色的概率为$\frac{\frac{1}{10}\times\frac{3}{4}}{\frac{3}{10}}=\frac{1}{4}=25\%$
此时是绿色的概率为75%，因此更可能是绿色

13.22

- a.
构建朴素贝叶斯模型的步骤如下：
    - 确定特征：选取单词作为文档分类的依据
    - 计算先验概率：对于给定的分好类的文档集，计算每类文档出现的概率$P(file)$，可以用该类文档数除以文档集总数得到
    - 计算条件概率：计算确定的单词在每类文档中出现的条件概率$P(word|file)$，可以用每类文档中，该单词的数量除以单词总数得到
    - 由此可得到朴素贝叶斯分类模型
- b.对于一个新文档，对其分类的过程如下：
    - 提取特征：提取新文档中出现的单词$word_1\dots word_n$

- 计算概率：根据确定的模型，计算新文档划分为不同类别的概率$P(file_i|newfile) \propto P(file_i) \times P(word_1)\dots P(word_n)$
- 确定类别：计算出的概率最大的类别就是新文档应该划分的类别

- c.题目给出的单词出现独立性显然不合理。因为很多单词是联合出现的，联合出现的概率比单独出现的概率要大得多，比如“United States”作为整体出现的概率比单独出现的概率要大得多，因此假设单词出现互相独立不合理

14.12

- (a).
  (ii)和(iii)都是正确表示。(i)是错误表示，因为N和F1、F2没有关系。(ii)是对题目的正确描述。(iii)是将M1、M2、N、F1、F2排序得到的网络，正确但更复杂。
- (b).
  由(a)可知(ii)最好。
- (c).
  分布表如下：

	N=1	N=2	N=3
M1=0	f+(1-f)e	f	f
M1=1	(1-f)(1-2e)	(1-f)e	0
M1=2	(1-f)e	(1-f)(1-2e)	(1-f)e
M1=3	0	(1-f)e	(1-f)(1-2e)
M1=4	0	0	(1-f)e

- (d).
  M1=1表明可能的N=0,1,2或N≥4。M2=3表明可能的N=2,3,4或N≥6。
  因此两个范围取交集即为可能结果，因此N可能为N=2,4或N≥6。
- (e).
  $P(N=2) \propto e \times e$
  $P(N=4) \propto f \times e$
  $P(N \geq 6) \propto f \times f$
  而e > f，因此N=2的可能性最大。

14.13

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