CatData HW3

Matthew Vanaman

03/11/19

A2.2, A2.5, A2.13, A2.21, A2.36. (3 points each)

A2.33 (5 points)

All code and work are shown in the appendix.

2.2

(a)

Answer:
Recoding this in a way that makes more sense to me. Will leave X and Y as they are:

- $X(0 = \text{no disease}, 1 = \text{disease})$
- $Y(0 = \text{negative}, 1 = \text{positive})$
- Sensitivity $= \pi_1 = P(Y = 1|X = 1) =$ probability positive diagnosis given disease
- Specificity $= 1 - \pi_2 = 1 - P(Y = 1|X = 0) = 1$ - probability positive diagnosis given no disease

After subtracting away the probability of positive diagnosis given no disease, you are left with probability of negative diagnosis given no disease:

$1 - P(Y = 1|X = 0) = P(Y = 0|X = 0) =$ probability negative diagnosis given no disease

Likewise, the "noise" of the test is captured by 1 - specificity. A good test would want to show that 1 - specificity yields a large probability as an effective test will have a large probability of correctly screening patients that do not have the disease.

(b)

Answer:
$P(X = 1|Y = 1) = \dfrac{P(Y = 1|X = 1)P(X = 1)}{P(Y = 1|X = 1)P(X = 1) + P(Y = 1|X = 0)P(X = 0)}$

Work:

$$\frac{\pi_1\gamma}{[\pi_1\gamma + \pi_2(1-\gamma)]} = \frac{P(B|A)P(A)}{P(B)} = \text{Bayes' rule}$$

In the above equation, P(B) represents P(X=1), or the probability of having the disease. But this is an unknown value expressed as γ, we have to use a particular version of Bayes' rule that takes into account the conditional probabilities associated with P(B):

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\widetilde{A})P(\widetilde{A})}.$$

Just replace with the problem's notation:

$$P(X = 1|Y = 1) = \frac{P(Y = 1|X = 1)P(X = 1)}{P(Y = 1|X = 1)P(X = 1) + P(Y = 1|X = 0)P(X = 0)}$$

(c)

Answer:
0.07

Work:

We need to solve for $P(X = 1|Y = 1)$. Lay out the pieces:

- $X(0 = \text{no disease}, 1 = \text{disease}.$
- $P(X = 1) = 0.01.$
- $Y(0 = \text{negative}, 1 = \text{positive}.$
- $\text{Sensitivity} = \pi_1 = P(Y = 1|X = 1) = 0.86.$
- $\text{Specificity} = 1 - \pi_2 = 1 - P(Y = 1|X = 0) = 0.88$
- If $1 - P(Y = 1|X = 0) = 0.88$, then $P(Y = 1|X = 0) = 0.12$.
- If $P(X = 1) = 0.01$, then $P(X = 0) = 1 - P(X = 1) = 0.99$.

Plug and chug:

$$P(X = 1|Y = 1) = \frac{P(Y = 1|X = 1)P(X = 1)}{P(Y = 1|X = 1)P(X = 1) + P(Y = 1|X = 0)P(X = 0)},$$

$$= \frac{0.86 \times 0.01}{(0.86 \times 0.01) + (0.12 \times 0.99)},$$

$$= 0.07.$$

```
# calculations
(0.86 * 0.01) / ((0.86 * 0.01) + (0.12 * 0.99))
```

```
[1] 0.06750392
```

(d)

Answer:

```
       neg    pos    ColMarg
no_dis 0.8514 0.1386 0.99
dis    0.0014 0.0086 0.01
```

Sensitivity (0.88) tells us probability of positive diagnosis when someone has the disease. If 0.01 proportion of 100 people have the disease, and the probability of being correctly diagnosed with the disease if you have it is 0.86, then 0.0086 proportion of 100 people will be correctly diagnosed as having the disease. Likewise, the specificity (0.86) gives us the probability of negative diagnosis with no disease. So out of proportion 0.99 of 100 people who do not have the disease, proportion 0.8514 out of 100 will be correctly diagnosed as not having the disease. Most of the people with the disease are in the positive diagnosis cell, which is a good thing. However, we can see based on the column margin that only about 7% of those who get diagnosed actually have the disease. Needless stress!

Work:
We start with what we know:
- Sensitivity $= \pi_1 = P(Y = 1|X = 1) = 0.86$.
- Specificity $= 1 - \pi_2 = 1 - P(Y = 1|X = 0) = 0.88$
- $P(Y = 1|X = 0) = 0.12$.
- $P(X = 0) = 0.99$.
- $P(X = 1) = 0.01$

```
       neg  pos  ColMarg
no_dis π_00 π_01 0.99
dis    π_10 π_11 0.01
```

Sensitivity (0.88) tells us probability of positive diagnosis when someone has the disease. If 0.01 proportion of 100 people have the disease, and the probability of being correctly diagnosed with the disease if you have it is 0.86, then 0.0086 proportion of 100 people will be correctly diagnosed as having the disease:

```r
0.01 * 0.86
```

```
[1] 0.0086
```

From there, because we have the margin and one of the cells, we have the other cell in that row:

```r
0.01 - 0.0086
```

```
[1] 0.0014
```

Likewise, the specificity (0.86) gives us the probability of negative diagnosis with no disease. So out of proportion 0.99 of 100 people who do not have the disease, proportion 0.8514 out of 100 will be correctly diagnosed as not having the disease:

```r
0.99 * 0.88
```

```
[1] 0.8712
```

Get the last cell:

```r
0.99 - 0.8514
```

```
[1] 0.1386
```

```
       neg    pos    ColMarg
no_dis 0.8514 0.1386 0.99
dis    0.0014 0.0086 0.01
```

```r
# check work
0.0086 + 0.0014
```

```
[1] 0.01
```

```r
0.1386 + 0.8514
```

```
[1] 0.99
```