Homework 4 UG

October 26, 2021

Jose Carlos Munoz

4.6)a

$$\begin{aligned}
P(\text{S}|\text{UG}) &= .15 \\
P(S|G) &= .23 \\
P(G) &= .2 \\
P(UG) &= .8
\end{aligned} \tag{1}$$

These are the known probabilites.
From this we can find $P(\text{G}|\text{S})$.
Because of Bayes Theroem $P(\text{G}|\text{S})$ is the same as the following

$$P(\text{G}|\text{S}) = \frac{P(\text{S}|\text{G}) * P(\text{G})}{P(\text{S})} \tag{1}$$

$P(\text{S})$ can be found as

$$\begin{aligned}
&P(\text{S}) = P(\text{S}|\text{G}) * P(\text{G}) + P(\text{S}|\text{UG}) * P(\text{UG}) \\
&P(\text{S}) = .23 * .2 + .15 * .8 \\
&P(\text{S}) = .166
\end{aligned} \tag{2}$$

Therefore

$$\begin{aligned}
P(\text{G}|\text{S}) &= \frac{.23 * .2}{.166} \\
P(\text{G}|\text{S}) &= .277
\end{aligned} \tag{3}$$

So the probabilty that a smoker is a graduate student is .277

4.6)c
The probability that a smoker is a graduated student can be written as $P(\text{UG}|\text{S})$.

$$\begin{aligned}
P(\text{UG}|\text{S}) &= \frac{P(\text{S}|\text{UG}) * P(\text{UG})}{P(\text{S})} \\
P(\text{UG}|\text{S}) &= \frac{.23 * .8}{.277} \\
P(\text{UG}|\text{S}) &= .857
\end{aligned} \tag{4}$$

So the probabilty that a smoker is an undergrad is .857.
Since $P(\text{UG}|\text{S}) > P(\text{G}|\text{S})$ we can conclude we have a higher chance of finding an undergrad that is a smoker

4.6)d

$$
\begin{aligned}
P(\text{D}|\text{UG}) &= .1 \\
P(\text{D}|\text{G}) &= .3 \\
P(\text{D}) &= P(\text{D}|\text{UG}) * P(\text{UG}) + P(\text{D}|\text{G}) * P(\text{G}) \\
P(\text{D}) &= 0.1 * .8 + .2 * .3 \\
P(\text{D}) &= .14 \\
P(\text{D,S}|\text{G}) &= P(\text{D}|\text{G}) * P(\text{S}|\text{G}) \\
P(\text{D,S}|\text{G}) &= .3 * .23 \\
P(\text{D,S}|\text{G}) &= .069 \\
P(\text{D,S}|\text{UG}) &= P(\text{D}|\text{UG}) * P(\text{S}|\text{UG}) \\
P(\text{D,S}|\text{UG}) &= .1 * .15 \\
P(\text{D,S}|\text{UG}) &= 0.015 \\
P(\text{D,S}) &= Q
\end{aligned} \tag{5}
$$

These are the known probabilites. Since we don't know what $P(\text{D,S})$ is, we set it as a constant Q
Now we can find the values for $P(\text{G}|\text{D,S})$ and $P(\text{UG}|\text{D,S})$

$$
\begin{aligned}
P(\text{UG}|\text{D,S}) &= \frac{P(\text{D,S}|\text{UG}) * P(\text{UG})}{P(\text{D,S})} \\
P(\text{UG}|\text{D,S}) &= \frac{.015 * .8}{Q} \\
P(\text{UG}|\text{D,S}) &= \frac{.012}{Q} \\
P(\text{G}|\text{D,S}) &= \frac{P(\text{D,S}|\text{UG}) * P(\text{UG})}{P(\text{D,S})} \\
P(\text{G}|\text{D,S}) &= \frac{.069 * .2}{Q} \\
P(\text{G}|\text{D,S}) &= \frac{.0139}{Q}
\end{aligned} \tag{6}
$$

From these results we can conclude that the chance that we find a graduate that lives in a dorm and is a smoker is is higher than the chance that we find an undergraduate that lives in a dorm and is a smoker.

4.7)a

$$
\begin{aligned}
P(\text{A=0}|+) &= \frac{2}{5} = .4 \\
P(\text{A=0}|-) &= \frac{3}{5} = .6 \\
P(\text{A=1}|+) &= \frac{3}{5} = .6 \\
P(\text{A=1}|-) &= \frac{2}{5} = .4 \\
P(\text{B=0}|+) &= \frac{4}{5} = .8 \\
P(\text{B=0}|-) &= \frac{3}{5} = .6 \\
P(\text{B=1}|+) &= \frac{1}{5} = .2 \\
P(\text{B=1}|-) &= \frac{2}{5} = .4 \\
P(\text{C=0}|+) &= \frac{3}{5} = .6 \\
P(\text{C=0}|-) &= \frac{0}{5} = 0 \\
P(\text{C=1}|+) &= \frac{2}{5} = .4 \\
P(\text{C=1}|-) &= \frac{5}{5} = .1
\end{aligned} \tag{7}
$$

4.7)b
we are task to find $P(\text{A=0,B=1,C=0}|+)$. Using the Bayes Therm we canfind the value as

$$
\begin{aligned}
P(+|\text{A=0,B=1,C=0}) &= \frac{P(\text{A=0,B=1,C=0}|+) * P(+)}{P(\text{A=0,B=1,C=0})} \\
P(+|\text{A=0,B=1,C=0}) &= \frac{P(\text{A=0}|+) * P(\text{B=1}|+) * P(\text{C=0}|+) * P(+)}{P(\text{A=0,B=1,C=0})} \\
P(+|\text{A=0,B=1,C=0}) &= \frac{.4 * .2 * .6 * .5}{P(\text{A=0,B=1,C=0})} \\
P(+|\text{A=0,B=1,C=0}) &= \frac{0.024}{P(\text{A=0,B=1,C=0})}
\end{aligned} \tag{8}
$$

$$
\begin{aligned}
P(\text{-}|\text{A=0,B=1,C=0}) &= \frac{P(\text{A=0,B=1,C=0|-}) * P(\text{-})}{P(\text{A=0,B=1,C=0})} \\
P(\text{-}|\text{A=0,B=1,C=0}) &= \frac{P(\text{A=0|-}) * P(\text{B=1|-}) * P(\text{C=0|-}) * P(\text{-})}{P(\text{A=0,B=1,C=0})} \\
P(\text{-}|\text{A=0,B=1,C=0}) &= \frac{.6 * .4 * 0 * .5}{P(\text{A=0,B=1,C=0})} \\
P(\text{-}|\text{A=0,B=1,C=0}) &= 0
\end{aligned} \tag{9}
$$

From these results we canconlude that the class label for (A=0, B=1, C=0) will be Class +.

4.7)c

We will be looking at the conditional probabilites for the them all over again with the m-estimate. When m=4 and p = 1/2; to find the new Conditonal probabilites we use this equation

$$\frac{n_c + m * p}{n + m} \tag{10}$$

so now the The conditional probablities will be

$$\begin{aligned}
P(\text{A=0}|+) &= \frac{2+2}{5+4} = \frac{4}{9} \\
P(\text{A=0}|-) &= \frac{3+2}{5+4} = \frac{5}{9} \\
P(\text{A=1}|+) &= \frac{3+2}{5+4} = \frac{5}{9} \\
P(\text{A=1}|-) &= \frac{2+2}{5+4} = \frac{4}{9} \\
P(\text{B=0}|+) &= \frac{4+2}{5+4} = \frac{6}{9} \\
P(\text{B=0}|-) &= \frac{3+2}{5+4} = \frac{5}{9} \\
P(\text{B=1}|+) &= \frac{1+2}{5+4} = \frac{3}{9} \\
P(\text{B=1}|-) &= \frac{2+2}{5+4} = \frac{4}{9} \\
P(\text{C=0}|+) &= \frac{3+2}{5+4} = \frac{5}{9} \\
P(\text{C=0}|-) &= \frac{0+2}{5+4} = \frac{2}{9} \\
P(\text{C=1}|+) &= \frac{2+2}{5+4} = \frac{4}{9} \\
P(\text{C=1}|-) &= \frac{5+2}{5+4} = \frac{7}{9}
\end{aligned} \tag{11}$$

4.7)d

we repeat b) but with the m-estimate conditional probabilities

$$\begin{aligned}
P(+|\text{A=0,B=1,C=0}) &= \frac{P(\text{A=0,B=1,C=0}|+) * P(+)}{P(\text{A=0,B=1,C=0})} \\
P(+|\text{A=0,B=1,C=0}) &= \frac{P(\text{A=0}|+) * P(\text{B=1}|+) * P(\text{C=0}|+) * P(+)}{P(\text{A=0,B=1,C=0})} \\
P(+|\text{A=0,B=1,C=0}) &= \frac{\frac{4}{9} * \frac{3}{9} * \frac{5}{9} * .5}{P(\text{A=0,B=1,C=0})} \\
P(+|\text{A=0,B=1,C=0}) &= \frac{0.0142}{P(\text{A=0,B=1,C=0})}
\end{aligned} \tag{12}$$

$$
\begin{aligned}
P(\text{-}|\text{A=0,B=1,C=0}) &= \frac{P(\text{A=0,B=1,C=0|-}) * P(\text{-})}{P(\text{A=0,B=1,C=0})} \\
P(\text{-}|\text{A=0,B=1,C=0}) &= \frac{P(\text{A=0|-}) * P(\text{B=1|-}) * P(\text{C=0|-}) * P(\text{-})}{P(\text{A=0,B=1,C=0})} \\
P(\text{-}|\text{A=0,B=1,C=0}) &= \frac{\frac{5}{9} * \frac{4}{9} * \frac{2}{9}.5}{P(\text{A=0,B=1,C=0})} \\
P(\text{-}|\text{A=0,B=1,C=0}) &= \frac{0.0274}{P(\text{A=0,B=1,C=0})}
\end{aligned} \tag{13}
$$

From these resutl we canconlude that the class label for (A=0,B=1,C=0) is class +

4.7)e

The better method would be the m-estimate becuase we do not want our entire expression to be zero