CS ID: huan2500

Question 3

Denote Gender as G, Prediction as P and Hyperlipidemia as H.

I used the formula below to calculate the probabilities in Excel:

P(P =YES | H =YES, G = Female) = $\frac{P(P=YES,H=YES,G=Female)}{P(P=YES,H=YES,G=Female)+P(P=NO,H=YES,G=Female)}$

P(P =YES | H =YES) = $\frac{\sum_G P(P=YES,H=YES)}{\sum_{P,G} P(H=YES)}$

P(H =YES | P =YES, G = Female) = $\frac{P(P=YES,H=YES,G=Female)}{P(P=YES,H=YES,G=Female)+P(P=YES,H=NO,G=Female)}$

P(H =YES | P =YES) = $\frac{\sum_P P(P=YES,H=YES)}{\sum_{H,G} P(P=YES)}$

Separated but not Sufficient

Table1		
G (Gender)	Female	Male
H (Hyperlipidemia)	YYYYYY NNN	YY NN
P (Prediction)	YYYNNN NNY	NY NN
Note: Y = YES, N = NO		
Values for P, H, G	Num of Instances	Probability
('YES', 'YES', 'Female')	3.00	0.23076923
('YES', 'YES', 'Male')	1.00	0.07692308
('YES', 'NO', 'Female')	1.00	0.07692308
('YES', 'NO', 'Male')	0.00	0
('NO', 'YES', 'Female')	3.00	0.23076923
('NO', 'YES', 'Male')	1.00	0.07692308
('NO', 'NO', 'Female')	2.00	0.15384615
('NO', 'NO', 'Male')	2.00	0.15384615
Sum	13.00	1
Separation		
P(P =YES | H =YES, G = Female)		0.50
P(P =YES | H =YES)		0.50
Sufficiency		
P(H =YES | P =YES, G = Female)		0.75
P(H =YES | P =YES)		0.80

Sufficient but not Separated

Table2		
G (Gender)	Female	Male
H (Hyperlipidemia)	YYYNNN NNY	NY NN
P (Prediction)	YYYYYY NNN	YY NN
Values for P, H, G	**Num of Instances**	**Probability**
('YES', 'YES', 'Female')	3.00	0.2307692
('YES', 'YES', 'Male')	1.00	0.0769231
('YES', 'NO', 'Female')	3.00	0.2307692
('YES', 'NO', 'Male')	1.00	0.0769231
('NO', 'YES', 'Female')	1.00	0.0769231
('NO', 'YES', 'Male')	0.00	0
('NO', 'NO', 'Female')	2.00	0.1538462
('NO', 'NO', 'Male')	2.00	0.1538462
Sum	13.00	1
Separation		
P(P =YES | H =YES, G = Female)		0.75
P(P =YES | H =YES)		0.80
Sufficiency		
P(H =YES | P =YES, G = Female)		0.50
P(H =YES | P =YES)		0.50

Conclusion

Separation and Sufficiency cannot hold together. For example, if we want to modify the values in Table 1 in order to hold sufficiency, the separation property would no longer hold any more. We cannot ensure both properties. Therefore, we can't enforce Sufficiency and Separation at the same time.

Note:

- Example details are shown in the two tables attached.
- Probability is calculated by $\frac{num\ of\ instances}{total\ instances}$ for each entry in the joint probability table.