CSC 384 A4 Part 3

Emilio Kartono

August 4, 2020

1. **Question 3 (10 points): Provide two examples to show that we can't enforce Sufficiency and Separation at the same time. More specifically, assume C is a classification, Y is a label representing 'ground truth', A is some 'protected attribute' (e.g. gender), and that all events in the joint distribution of (A,C,Y) have positive probability. Define Separation as meaning A is independent of C given Y, and Sufficiency as meaning A is independent of Y given C. Give one example where sufficiency holds but not separation, and one where separation holds but not sufficiency. Illustrate your examples with joint probability tables for both situations (i.e. P(A,C,Y) for all combinations of value assignments to variables), so that we can verify the conditional independence relations.**

   Let separation = $P(C|Y,A) = P(C|Y)$
   Let sufficiency = $P(Y|C,A) = P(Y|C)$
   We want to show two examples where separation makes sufficiency not hold, and sufficiency makes separation not hold.

   (a) An example that separation holds $\implies$ sufficiency does not hold
   Let C = BMI
   Let Y = Hyperlipidemia
   Let A = Gender

   Note that $P(C|Y,A) = P(C|Y)$ because for $P(C|Y,A)$, we get this table:

C	Y	A	$P(C\|Y,A)$
< 18.5	YES	$Male$	0.003683854066263697
< 18.5	YES	$Female$	0.003683854066263697
< 18.5	NO	$Male$	0.026066985641622183
< 18.5	NO	$Female$	0.026066985641622183
18.5	YES	$Male$	0.198529742653345
18.5	YES	$Female$	0.198529742653345
18.5	NO	$Male$	0.49179708484152856
18.5	NO	$Female$	0.49179708484152856
24.0	YES	$Male$	0.46963281260011486
24.0	YES	$Female$	0.46963281260011486
24.0	NO	$Male$	0.3626723133666205
24.0	NO	$Female$	0.3626723133666205
28.0	YES	$Male$	0.3281535906802762
28.0	YES	$Female$	0.3281535906802762
28.0	NO	$Male$	0.11946361615022867
28.0	NO	$Female$	0.11946361615022867

Moreover, for $P(C|Y)$, we get this table:

C	Y	$P(C\|Y)$
< 18.5	YES	0.003683854066263697
< 18.5	NO	0.026066985641622183
18.5	YES	0.198529742653345
18.5	NO	0.49179708484152856
24.0	YES	0.46963281260011486
24.0	NO	0.3626723133666205
28.0	YES	0.3281535906802762
28.0	NO	0.11946361615022867

Thus, $\forall c \in C.domain(), \forall y \in Y.domain(), \forall a \in A.domain(), P(C|Y, A) = P(C|Y)) \implies$ separation.
However, sufficiency does not hold because for $P(Y|C, A)$, we get this table:

Y	C	A	$P(Y\|C, A)$
YES	18.5	$Male$	0.24110864246533822
NO	18.5	$Male$	0.7588913575346617
YES	18.5	$Female$	0.18899641307583723
NO	18.5	$Female$	0.8110035869241627
YES	24.0	$Male$	0.5047417852429753
NO	24.0	$Male$	0.49525821475702475
YES	24.0	$Female$	0.4277674560469438
NO	24.0	$Female$	0.5722325439530562
YES	28.0	$Male$	0.6837341865229888
NO	28.0	$Male$	0.31626581347701127
YES	28.0	$Female$	0.6132638858790895
NO	28.0	$Female$	0.3867361141209104
YES	< 18.5	$Male$	0.10009280710767367
NO	< 18.5	$Male$	0.8999071928923262
YES	< 18.5	$Female$	0.07542975141421693
NO	< 18.5	$Female$	0.924570248585783

Thus, $y \in Y c \in C s.t. P(Y|C, A = \text{Male})! = P(Y|C, A = \text{Female}) \implies$ sufficiency does not hold.

(b) An example that sufficiency holds $\implies$ separation does not hold:
Let A = Gender Let C = Hyperlipidemia Let Y = Central Obesity
Then, $P(Y|C, A) =$

Y	C	A	$P(Y\|C, A)$
YES	YES	$Male$	0.7876943228020956
NO	YES	$Male$	0.21230567719790436
YES	YES	$Female$	0.7876943228020957
NO	YES	$Female$	0.21230567719790436
YES	NO	$Male$	0.583203690852557
NO	NO	$Male$	0.416796309147443
YES	NO	$Female$	0.583203690852557
NO	NO	$Female$	0.41679630914744303

Also, $P(Y|C) =$

Y	C	$P(Y\|C)$
YES	YES	0.7876943228020957
NO	YES	0.21230567719790436
YES	NO	0.583203690852557
NO	NO	0.41679630914744303

Thus, $\forall a \in A.domain(), \forall c \in C.domain(), \forall y \in Y.domain(), P(A|C,Y) = P(A|C) \implies$ sufficiency.

However, for $P(C|A,Y) =$

C	A	Y	$P(C\|A,Y)$
YES	$Male$	YES	0.5152677192429572
NO	$Male$	YES	0.4847322807570428
YES	$Female$	YES	0.4381081553691644
NO	$Female$	YES	0.5618918446308354
YES	$Male$	NO	0.2861709298888966
NO	$Male$	NO	0.7138290701111034
YES	$Female$	NO	0.22723545270995207
NO	$Female$	NO	0.7727645472900478

Note that $P(Hyperlipidemia = yes|Gender = male, CentralObesity = yes) \neq P(Hyperlipidemia = yes|Gender = female, CentralObesity = yes)$.

Thus, for A = Gender, C = Hyperlipidemia, and Y = Central Obesity, sufficiency holds but separation does not.

Thus, from the two examples, we cannot enforce Sufficiency and Separation at the same time.