CSC384 a4

Jiakai Shi

Student number: 1003986760

1 Question 3

My Bayes Network is Fair. Here is a situation where "Separated" holds but not "Sufficient":

- Firstly, it satisfies "Separated" property.
  To prove this, we need to know whether the predictions(CentralObesity = YES) are well 'Separated' from gender, meaning:

$$P(CentralObesity = YES \| Hyperlipidemia = YES, Gender = Female) = P(CentralObesity = YES \| Hyperlipidemia = YES)$$

  Here is the modified code in medicalDiagnosis.py that can test this:

```
a = gd
c = co
y = hl

print("Prediction: ", c.name)
print()
y.set_evidence('YES')
a.set_evidence('Female')
probs = VE(medical, c, [y, a])
probs1 = VE(medical, c, [y])
doms = c.domain()
for i in range(len(probs)):
    txt = "P({V_0:} = {d_0:} | {V_1:} = {d_1:},\
    {V_2:} = {d_2:}) = {probability:0.1f}"
    print(txt.format(V_0 = c.name, d_0 = doms[i],\
    V_1 = y.name, d_1 = y.get_evidence(), V_2 =\
    a.name, d_2 = a.get_evidence(), probability = 100*probs[i]))
    print()
for j in range(len(probs1)):
    txt = "P({V_0:} = {d_0:} | {V_1:} = {d_1:})\
    = {probability:0.1f}"
    print(txt.format(V_0 = c.name, d_0 = doms[j],\
    V_1 = y.name, d_1 = y.get_evidence(), \
    probability = 100*probs1[j]))
    print()
```

And here is the output:

```
Prediction:   CentralObesity

P(CentralObesity = YES | Hyperlipidemia = YES, Gender = Female) = 78.8

P(CentralObesity = NO | Hyperlipidemia = YES, Gender = Female) = 21.2

P(CentralObesity = YES | Hyperlipidemia = YES) = 78.8

P(CentralObesity = NO | Hyperlipidemia = YES) = 21.2
```

We can see that in the assignment CentralObesity = YES,
P(CentralObesity = YES || Hyperlipidemia = YES, Gender = Female) =
P(CentralObesity = YES || Hyperlipidemia = YES) = 78.8, thus my Bayes Network satisfies 'Separated' property.

- Secondly, it does not satisfy 'Sufficient' property.
  To prove this, we need to know whether the predictions(CentralObesity = YES) are not 'Sufficient'(and gender tells us nothing more than our label about the presence of disease), meaning:

$$P(Hyperlipidemia = YES \| CentralObesity = YES, Gender = Female) \neq P(Hyperlipidemia = YES \| CentralObesity = YES)$$

  Here is the modified code in medicalDiagnosis.py that can test this:

```
a = gd
c = co
y = hl

print("Prediction: ", c.name)
print()
for dom in c.domain():
    c.set_evidence(dom)
    a.set_evidence('Female')
    probs = VE(medical, y, [c, a])
    probs1 = VE(medical, y, [c])
    doms = y.domain()
    for i in range(len(probs)):
        txt = "P({V_0:} = {d_0:} | {V_1:} = {d_1:}, \
        {V_2:} = {d_2:}) = {probability:0.1f}"
        print(txt.format(V_0 = y.name, d_0 = doms[i], \
        V_1 = c.name, d_1 = c.get_evidence(), V_2 = a.name,\
        d_2 = a.get_evidence(), probability = 100*probs[i]))
        print()
    for j in range(len(probs1)):
        txt = "P({V_0:} = {d_0:} | {V_1:} = {d_1:})\
       = {probability:0.1f}"
```

```
print(txt.format(V_0 = y.name, d_0 = doms[j],\
V_1 = c.name, d_1 = c.get_evidence(), \
probability = 100*probs1[j]))
print()
```

And here is the output:

```
Prediction:  CentralObesity

P(Hyperlipidemia = YES | CentralObesity = YES, Gender = Female) = 43.8

P(Hyperlipidemia = NO | CentralObesity = YES, Gender = Female) = 56.2

P(Hyperlipidemia = YES | CentralObesity = YES) = 47.9

P(Hyperlipidemia = NO | CentralObesity = YES) = 52.1

P(Hyperlipidemia = YES | CentralObesity = NO, Gender = Female) = 22.7

P(Hyperlipidemia = NO | CentralObesity = NO, Gender = Female) = 77.3

P(Hyperlipidemia = YES | CentralObesity = NO) = 25.8

P(Hyperlipidemia = NO | CentralObesity = NO) = 74.2
```

We can see that in the assignment Hyperlipidemia = YES, given evidence CentralObesity = YES and Gender = Female, P(Hyperlipidemia = YES — CentralObesity = YES, Gender = Female) = 43.8, however, P(Hyperlipidemia = YES — CentralObesity = YES) = 47.9, thus it does not satisfy "Sufficient" property while it satisfies "Separated" property.

Here is an alternative situation where "Sufficiency" holds but not "Separation":

- Firstly, it satisfies "Sufficiency" property.
  To prove this, we need to know whether the predictions(Hyperlipidemia = YES) are "Sufficient"(and gender tells us nothing more than our label about the presence of disease), meaning:

$$P(BMI = \ 18.5 \| Hyperlipidemia = YES, Gender = Female) = \\ P(BMI = \ 18.5 \| Hyperlipidemia = YES)$$

  And we only change the assignment of c and a, in this situation: $A = Gender, C = Hyperlipidemia, Y = BMI$, and the rest of code stays unchanged.
  And here is the output:

```
Prediction:  Hyperlipidemia

P(BMI = ~18.5 | Hyperlipidemia = YES, Gender = Female) = 19.9

P(BMI = ~24.0 | Hyperlipidemia = YES, Gender = Female) = 47.0
```

```
P(BMI = ~28.0 | Hyperlipidemia = YES, Gender = Female) = 32.8

P(BMI = <18.5 | Hyperlipidemia = YES, Gender = Female) = 0.4

P(BMI = ~18.5 | Hyperlipidemia = YES) = 19.9

P(BMI = ~24.0 | Hyperlipidemia = YES) = 47.0

P(BMI = ~28.0 | Hyperlipidemia = YES) = 32.8

P(BMI = <18.5 | Hyperlipidemia = YES) = 0.4

P(BMI = ~18.5 | Hyperlipidemia = NO, Gender = Female) = 49.2

P(BMI = ~24.0 | Hyperlipidemia = NO, Gender = Female) = 36.3

P(BMI = ~28.0 | Hyperlipidemia = NO, Gender = Female) = 11.9

P(BMI = <18.5 | Hyperlipidemia = NO, Gender = Female) = 2.6

P(BMI = ~18.5 | Hyperlipidemia = NO) = 49.2

P(BMI = ~24.0 | Hyperlipidemia = NO) = 36.3

P(BMI = ~28.0 | Hyperlipidemia = NO) = 11.9

P(BMI = <18.5 | Hyperlipidemia = NO) = 2.6
```

We can observe that in the assignment BMI =~18.5, P(BMI =~18.5 || Hyperlipidemia = YES, Gender = Female) = P(BMI =~18.5 || Hyperlipidemia = YES) = 19.9, which suggests that it satisfies "Sufficient" property.

- Secondly, it does not satisfy 'Separate' property.
  To prove this, we need to know whether the predictions(Hyperlipidemia = YES) are not 'Separated' from gender, meaning:

$$P(Hyperlipidemia = YES \| BMI = \ 18.5, Gender = Female) \neq$$
$$P(Hyperlipidemia = YES \| BMI = \ 18.5)$$

  And we only change the assignment of c and a, in this situation: $A = Gender, C = Hyperlipidemia, Y = BMI$, and the rest of code stays unchanged.
  And here is the output:

```
Prediction:  Hyperlipidemia

P(Hyperlipidemia = YES | BMI = ~18.5, Gender = Female) = 18.9
```

```
P(Hyperlipidemia = NO | BMI = ~18.5, Gender = Female) = 81.1

P(Hyperlipidemia = YES | BMI = ~18.5) = 21.6

P(Hyperlipidemia = NO | BMI = ~18.5) = 78.4
```

We can see that in the assignment Hyperlipidemia = YES, given evidence BMI =˜ 18.5 and Gender = Female, P(Hyperlipidemia = YES ‖ BMI =˜ 18.5, Gender = Female) = 18.9, however, P(Hyperlipidemia = YES ‖ BMI =˜ 18.5) = 21.6, thus it does not satisfy ”Separated” property while it satisfies ”Sufficient” property.

Therefore, my Bayes Network is Fair.