MACHINE LEARNING

HOMEWORK PROBABILISTIC GRAPHIC MODELS

ELİF HANGÜL

1. From the textbook of C. Bishop: 8.3, 8.4

8.3)

a	b	c	$p(a,b,c)$
0	0	0	0.192
0	0	1	0.144
0	1	0	0.048
0	1	1	0.216
1	0	0	0.192
1	0	1	0.064
1	1	0	0.048
1	1	1	0.096

8.3 ($\star\star$) Consider three binary variables $a, b, c \in \{0, 1\}$ having the joint distribution given in Table 8.2. Show by direct evaluation that this distribution has the property that a and b are marginally dependent, so that $p(a, b) \neq p(a)p(b)$, but that they become independent when conditioned on c, so that $p(a, b|c) = p(a|c)p(b|c)$ for both $c = 0$ and $c = 1$.

Answer:
Based on definition, we can write the below values:

$$p(a,b) = p(a,b, c = 0) + p(a,b, c = 1) = \begin{cases} 0.192 + 1.144 = 0.336, \text{if } a = 0, b = 0 \\ 0.048 + 0.216 = 0.264, \text{if } a = 0, b = 1 \\ 0.192 + 0.064 = 0.256, \text{if } a = 1, b = 0 \\ 0.048 + 0.096 = 0.144, \text{if } a = 1, b = 1 \end{cases}$$

Similarly, we can calculate the p(a) and p(b) as:

$$p(a) = p(a,b = 0) + p(a,b = 1) = \begin{cases} 0.192 + 0.144 + 0.048 + 0.216 = 0.6, \text{if } a = 0 \\ 0.192 + 0.064 + 0.048 + 0.096 = 0.4, \text{if } a = 1 \end{cases}$$

$$p(b) = p(a = 0,b) + p(a = 1,b) = \begin{cases} 0.192 + 0.144 + 0.192 + 0.064 = 0.592, \text{if } b = 0 \\ 0.048 + 0.216 + 0.048 + 0.096 = 0.408, \text{if } b = 1 \end{cases}$$

In conclusion we can see that

$p(a,b) \neq p(a)p(b)$.

Let’s see this non equality with an example.

For instance we have

$p(a=0,b=1) = 0.264$

p(a=0) = 0.6

p(b=1) = 0.408

p(a)p(b) = 0.6 x 0.408 = 0.2448

Since 0.264≠0.2448, we can show that p(a=0,b=1)≠p(a=0)p(b=1)

- a and b are marginally dependent.

Now let's show that a and b become independent when conditioned on c.

First calculate p(c):

$$p(c) = \sum_{a,b=0,1} p(a,b,c) = \begin{cases} 0.192 + 0.048 + 0.192 + 0.048 = 0.480 \; if \; c = 0 \\ 0.144 + 0.216 + 0.064 + 0.096 = 0.520 \text{ if c } = 1 \end{cases}$$

According to Bayes' Theorem, we have:

$$\text{p(a,b|c)} = \frac{p(a,b,c)}{p(c)} = \begin{cases} \frac{0.192}{0.480} = 0.400, \text{if a } = 0, \text{b } = 0, \text{c } = 0 \\ \frac{0.144}{0.520} = 0.277, \text{if a } = 0, \text{b } = 0, \text{c } = 1 \\ \frac{0.048}{0.480} = 0.100, \text{if a } = 0, \text{b } = 1, \text{c } = 0 \\ \frac{0.216}{0.520} = 0.415, \text{if a } = 0, \text{b } = 1, \text{c } = 1 \\ \frac{0.192}{0.480} = 0.400, \text{if a } = 1, \text{b } = 0, \text{c } = 0 \\ \frac{0.064}{0.520} = 0.123, \text{if a } = 1, \text{b } = 0, \text{c } = 1 \\ \frac{0.048}{0.480} = 0.100, \text{if a } = 1, \text{b } = 1, \text{c } = 0 \\ \frac{0.096}{0.520} = 0.185, \text{if a } = 1, \text{b } = 1, \text{c } = 1 \end{cases}$$

Now let's calculate the p(a,c) and p(b,c)

$$p(a,c) = p(a, b = 0, c) + p(a, b = 1, c) = \begin{cases} 0.192 + 0.048 = 0.240, if \; a = 0, c = 0 \\ 0.144 + 0.216 = 0.360, if \; a = 0, c = 1 \\ 0.192 + 0.048 = 0.240, if \; a = 1, c = 0 \\ 0.064 + 0.096 = 0.160, if \; a = 1, c = 1 \end{cases}$$

$$p(b,c) = p(a = 0, b, c) + p(a = 1, b, c) = \begin{cases} 0.192 + 0.192 = 0.384, if \; b = 0, c = 0 \\ 0.144 + 0.064 = 0.208, if \; b = 0, c = 1 \\ 0.048 + 0.048 = 0.096, if \; b = 1, c = 0 \\ 0.216 + 0.096 = 0.312, if \; b = 1, c = 1 \end{cases}$$

Now we have every value to calculate p(a|c) and p(b|c):

$$p(a|c) = \frac{p(a,c)}{p(c)} = \begin{cases} 0.240/0.480 = 0.500, \text{if a} = 0, \text{c} = 0 \\ 0.360/0.520 = 0.692, \text{if a} = 0, \text{c} = 1 \\ 0.240/0.480 = 0.500, \text{if a} = 1, \text{c} = 0 \\ 0.160/0.520 = 0.308, \text{if a} = 1, \text{c} = 1 \end{cases}$$

$$p(b|c) = \frac{p(b,c)}{p(c)} = \begin{cases} 0.384/0.480 = 0.800, \text{if b} = 0, \text{c} = 0 \\ 0.208/0.520 = 0.400, \text{if b} = 0, \text{c} = 1 \\ 0.096/0.480 = 0.200, \text{if b} = 1, \text{c} = 0 \\ 0.312/0.520 = 0.600, \text{if b} = 1, \text{c} = 1 \end{cases}$$

Now we can easily verify the statement

p(a,b|c) = p(a|c)p(b|c).

Let's see this equality with an example.

For instance we have:

p(a = 0,b = 1|c = 0) = 0.1

p(a = 0|c = 0) = 0.500

p(b = 1|c = 0) = 0.200

0.1 = 0.5 x 0.2

p(a=0,b=1|c=0) = p(a=0|c=0)p(b=1|c=0).

- a and b are independent when conditioned on c

8.4 ($\star\star$) Evaluate the distributions $p(a)$, $p(b|c)$, and $p(c|a)$ corresponding to the joint distribution given in Table 8.2. Hence show by direct evaluation that $p(a, b, c) = p(a)p(c|a)p(b|c)$. Draw the corresponding directed graph.

Answer:

We already calculated the p(a) and p(b|c) values from the question 8.3, let's rewrite them here.

p(a) = p(a,b = 0)+ p(a,b = 1) = $\begin{cases} 0.6, \text{if a} = 0 \\ 0.4, \text{if a} = 1 \end{cases}$

And

$$p(b|c) = \frac{p(b,c)}{p(c)} = \begin{cases} 0.384/0.480 = 0.800, \text{if b} = 0, \text{c} = 0 \\ 0.208/0.520 = 0.400, \text{if b} = 0, \text{c} = 1 \\ 0.096/0.480 = 0.200, \text{if b} = 1, \text{c} = 0 \\ 0.312/0.520 = 0.600, \text{if b} = 1, \text{c} = 1 \end{cases}$$

We can also obtain p(c|a):

$$p(c|a) = \frac{p(a,c)}{p(a)} = \begin{cases} 0.24/0.6 = 0.4, \text{if a} = 0, \text{c} = 0 \\ 0.36/0.6 = 0.6, \text{if a} = 0, \text{c} = 1 \\ 0.24/0.4 = 0.6, \text{if a} = 1, \text{c} = 0 \\ 0.16/0.4 = 0.4, \text{if a} = 1, \text{c} = 1 \end{cases}$$

Now we can easily verify the statement that p(a,b, c) = p(a)p(c|a)p(b|c) according to the values that we calculated.

Let's give an example.

p(a=0,b=1,c=1) = 0.216

p(a=0) = 0.6

p(c=1|a=0) = 0.6

p(b=1|c=1) = 0.6

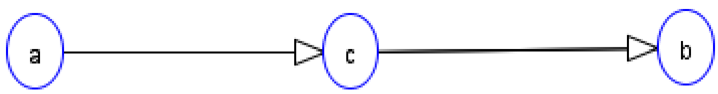
0.216 = 0.6 * 0.6 * 0.6

p(a=0,b=1,c=1) = p(a=0)p(c=1|a=0)p(b=1|c=1)

⇨ Similarly for every p(a,b,c) the equality of p(a,b, c) = p(a)p(c|a)p(b|c) can be shown.

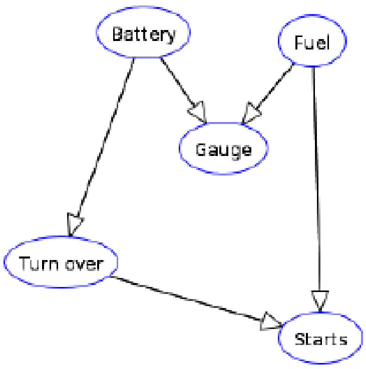
The directed graph looks like:

a ⟶ c ⟶ b



2. The Bayesian network shown in the figure captures the probability of a car engine starting.

- The Battery (B) of the car can be either in *good* or *bad* condition.
- The Fuel tank (F) can be either *empty* or *not empty*.
- The Gauge or indicator (G) can also indicate *empty* or *not empty*.
- The turn over (T) can be *false* or *true*.
- The engine starts (S) can be *yes* or *no*.



(a) Write down the factorization of the joint probability p(B; F; G; T; S) induced by the network.

Answer: P(B,F,G,T,S) = P(B)P(F)P(G|F,B)P(T|B)P(S|F,T)

(b) Argue whether the following conditional independences are satisfied or not.

i. Is T independent of F if no evidence is provided?

Answer: We look all the paths from T to F to see if all the paths are blocked or not. There are two paths from T to F.

Path 1) Turn Over =>Starts<=Fuel. With no evidence this head to head model becomes a blocked path.

Path 2) Has two parts:
Part 1= (Turn Over<=Battery=>Gauge)
Part 2= (Battery=>Gauge<=Fuel)

Since to get to Fuel from Turn Over we need to look the combined parts. (Part 1+Part2)

Part 1 is a tail to tail model and with no evidence the path is not blocked.
Part 2 is a head to head model and with no evidence the path is blocked.

In general, since a part of it blocked, we cannot reach the Fuel from Turn Over so we can say path 2 is blocked.

- Since both Path1 and Path2 are blocked then we can say T is independent of F given no evidence.

ii. Is T independent of F if we observe that the engine does not start?

Answer: We essentially look into the paths from the previous question. Path 1(Turn Over =>Starts<=Fuel) becomes a head to head model with evidence becomes dependent. The Path 2 still independent but since we have a dependent path we cannot say all the paths are blocked so T and F becomes dependent given an evidence for S.

- T is not independent of F given S.

iii. Is B independent of F if we observe that the engine starts?

Answer: From B to F, there are two paths.

Path 1) Battery =>Gauge<=Fuel. With no evidence, this head to head model becomes independent. The path is blocked.
Path2) Has two parts:
Part 1) Battery=>Turn Over =>Starts
Part 2) Turn Over=>Starts<=Fuel

Part 1 is a head to tail model with no evidence, so this path is not blocked.
Part 2 is a head to head model with evidence, so this path is not blocked.

Since Path 2 is combination of Part1 and Part 2 (Path 2 = Part1&Part2), this path is not blocked.

- Since not all paths are blocked, B is not independent of F given S.

(c) The following conditional probability tables fully define the model:

- P(B = bad) = 0:02
- P(F = empty) = 0:05
- P(G = empty|B = good; F = notempty) = 0:04
- P(G = empty|B = good; F = empty) = 0:97
- P(G = empty|B = bad; F = notempty) = 0:1
- P(G = empty|B = bad; F = empty) = 0:99
- P(T = false|B = good) = 0:03
- P(T = false|B = bad) = 0:98
- P(S = no|T = true; F = notempty) = 0:01
- P(S = no|T = false; F = notempty) = 1
- P(S = no|T = true; F = empty) = 0:92
- P(S = no|T = false; F = empty) = 0:99

We observe the car does not start (S = no). Calculate the probability that the fuel tank is empty.
P(F = empty|S = no). (Try to use the least possible number of terms in your calculations).

Answer:

I used T and F variables as domain for the nodes. The denotations are:

B: good=T, bad=F

F: nonempty=T, empty=F

G: not empty=T, empty=F

Turn Over: T or F

S: yes: T, no: F

We need to calculate the p(F = empty|S = no) which is p(Fuel=F|Starts=F) from our denotations.

To calculate p(F=empty|S=no) let's define some probability functions.

$$f_0 = p(Battery)$$

$$f_0(Battery = F) = 0.02 \quad f_0(Battery = T) = 0.98$$

$$f_1 = p(Battery, Turn\ Over)$$

$$f_1(Battery = F, Turn\ Over = T) = 0.02$$

$$f_1P(Battery = T, Turn\ Over = T) = 0.97$$

$The\ values\ for\ f_0\ and\ f_1$ are already given in the question.

Let f_2 defined as p(Turn Over).

Now calculation the $f_2(Turn\ Over = T)\ and\ f_2(Turn\ Over = F)$ values:

$$f_2(Turn\ Over = T) = f_0(Battery = T) * f_1(Battery = T, Turn\ Over = T) + f_0(Battery = F) * f_1(Battery = F, Turn\ Over = T)$$

= (0.98 * 0.97) + (0.02 * 0.02) = 0.951

$$f_2(Turn\ Over = F) = 1 - f_2(Turn\ Over = T) = 0.049$$

Let f_3 defined as p(Fuel, Turn Over)

Now calculation the $f_3(Fuel = F, Turn\ Over = T), f_3(Fuel = F, Turn\ Over = F), f_3(Fuel = T, Turn\ Over = T), f_3(Fuel = T, Turn\ Over = T)$ values which are already given in the question:

$$f_3(Fuel = F, Turn\ Over = T) = p(Starts = F|Fuel = F, Turn\ Over = T) = 0.92$$

$$f_3(Fuel = F, Turn\ Over = F) = p(Starts = F|Fuel = F, Turn\ Over = F) = 0.99$$

$$f_3(Fuel = T, Turn\ Over = T) = p(Starts = F|Fuel = T, Turn\ Over = T) = 0.01$$

$$f_3(Fuel = T, Turn\ Over = F) = p(Starts = F|Fuel = T, Turn\ Over = F) = 1.0$$

Let f_4 defined as p(Fuel)

Now calculation the $f_4(Fuel = F)$ value:

$$f_4(Fuel = F) = f_3(Fuel = F, Turn\ Over = T) * f_2(Turn\ Over = T) + f_3(Fuel = F, Turn\ Over = F) * f_2(Turn\ Over = F)$$

= (0.92 * 0.951) + (0.99 * 0.049) = 0.92343

Let f_5 denote the prior priority of Fuel. f_5 (Fuel=empty) which is given as 0.05

p(F = empty,S = no) = p(Fuel=F,Starts=F) = $f_4(Fuel = F) * \ f_5(Fuel = F)$

= 0.92343 * 0.05 = 0.04617

Since we are looking for p(Fuel=empty|S=no) we also need to calculate p(Fuel=nonempty|S=no)

Now calculation the $f_4(Fuel = T)$ value:

$$f_4(Fuel = T) = \ f_3(Fuel = T, Turn\ Over = T) * f_2(Turn\ Over = T) \\ + \ f_3(Fuel = T, Turn\ Over = F) * f_2(Turn\ Over = F)$$

= (0.01 * 0.951) + (1.0 * 0.049) = 0.05851

$The\ prior\ probability\ of\ f_5$ (Fuel=nonempty) is given as 0.95

p(F = nonempty,S = no) = p(Fuel=T,Starts=F) = $f_4(Fuel = T) * \ f_5(Fuel = T)$

= 0.05851 * 0.95 = 0.0555845

p(Starts=no) = p(Fuel=nonempty, S=no) + p(Fuel=empty, S=no) which can be written as with our denotations:

p(Starts=F) = p(Fuel=T, Starts=F) + p(Fuel=F, Starts=F) = 0.0555845 + 0.04617 = 0.1017545

- $\mathrm{p(Fuel = empty|S = no)} = \frac{\mathrm{p(F = empty, S = no)}}{\mathrm{p(S=no)}} = \frac{0.04617}{0.1017545}$ = 0.4537

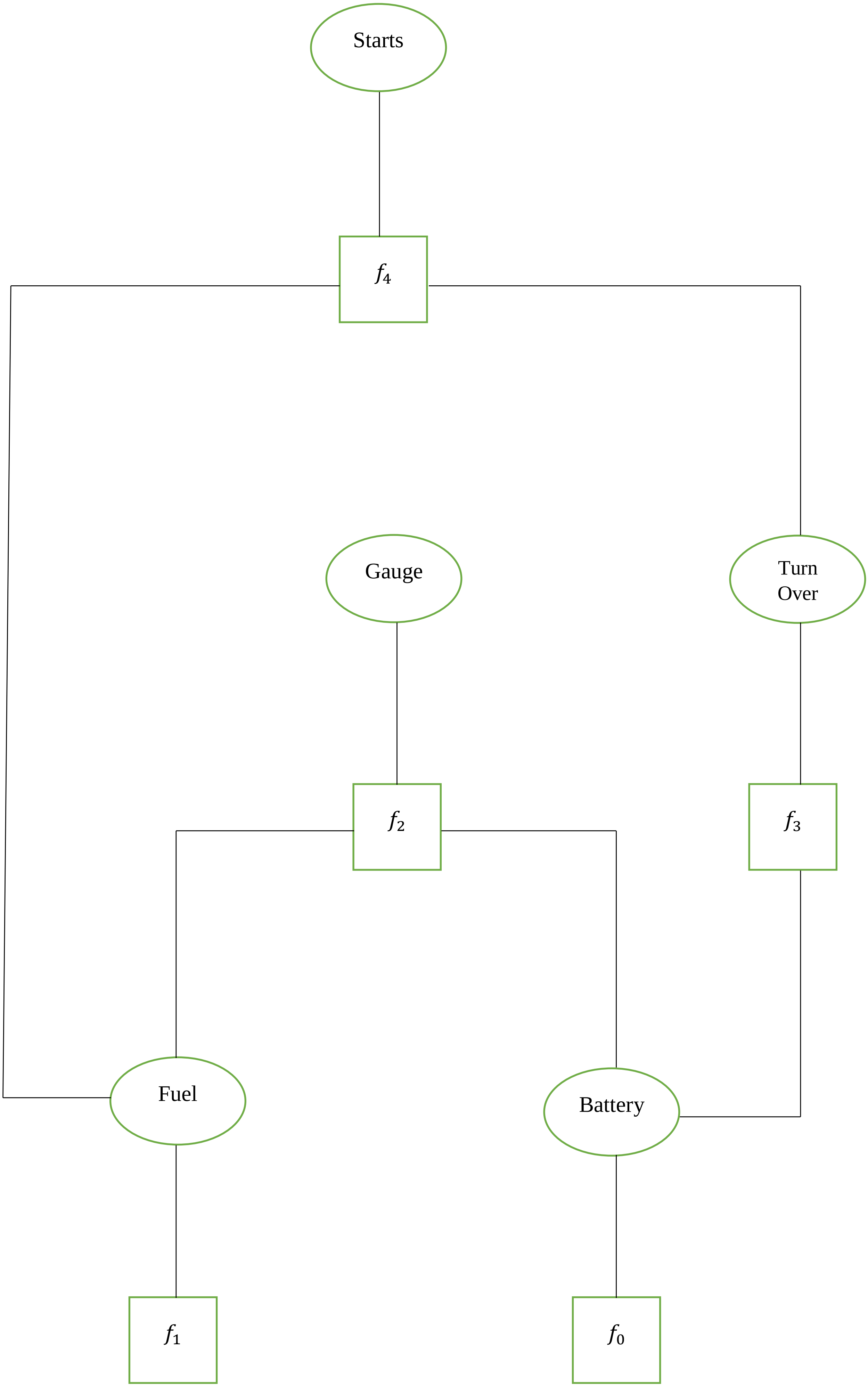
⇨ The results are also verified by Java tool http://aispace.org/bayes/

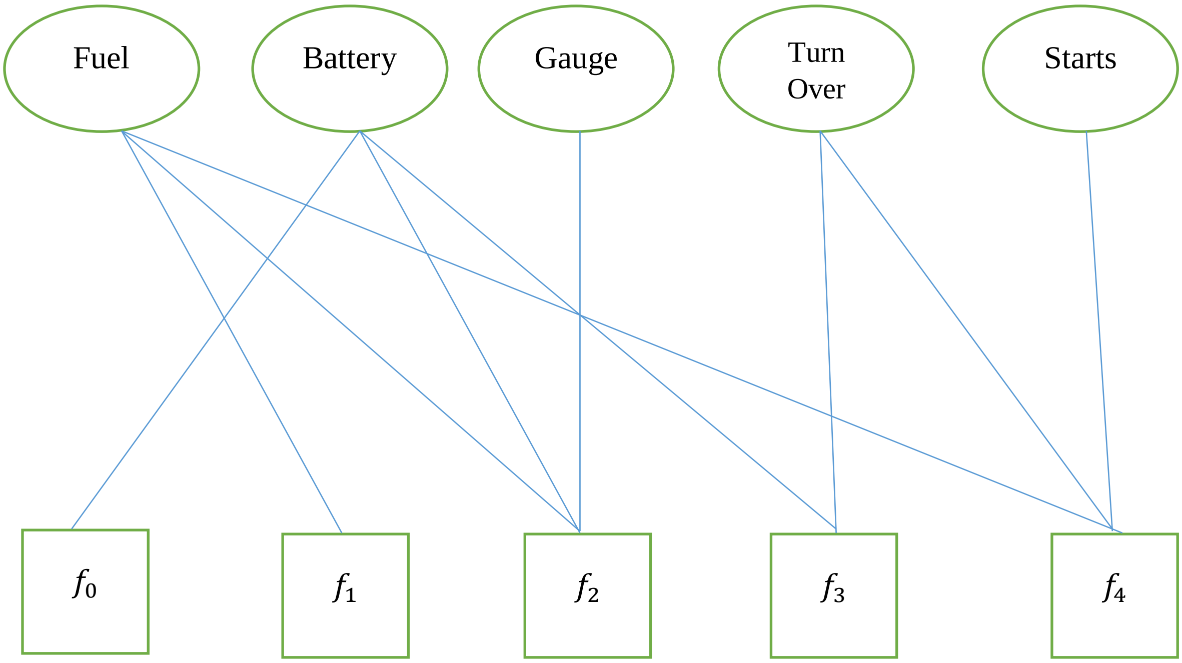
(d) Transform the model into an equivalent factor graph and describe it fully, e.g., write down all the involved factors and their values.

Answer:

Let’s define the factors as:

- ❖ $f_0(Battery)$
- ❖ $f_1(Fuel)$
- ❖ $f_2(Gauge|Battery, Fuel)$
- ❖ $f_3(Turn\ Over|Battery)$
- ❖ $f_4(Starts|Fuel, Turn\ Over)$


Starts
f_4
Gauge
Turn Over
f_2
f_3
Fuel
Battery
f_1
f_0



The values of f’s based on the values of B,G,S,T,F are shown in the tables below.

B	f_0
0	0.02
1	0.98

F	f_1
0	0.05
1	0.95

B	F	G	f_2
1	1	1	0.96
1	1	0	0.04
1	0	1	0.03
1	0	0	0.97
0	1	1	0.9
0	1	0	0.1
0	0	1	0.01
0	0	0	0.99

B	T	f_3
1	1	0.97
1	0	0.03
0	1	0.02
0	0	0.98

F	T	S	f_4
1	1	1	0.99
1	1	0	0.01
1	0	1	0.0
1	0	0	1.0
0	1	1	0.08
0	1	0	0.92
0	0	1	0.01
0	0	0	0.99

(e) Write down the belief propagation messages from variables B; F to variable G (You can assume other required messages for these calculations equal to ones).

Answer:

Rule of the belief propagation message algorithm is to take the product over all incoming messages except the message from the outgoing edge for all variable nodes.

According to this let's write messages for B.

From f_0 and f_3 there are incoming messages to B

From B to f_2 there is an outgoing message which reaches to G

The outgoing message could be written as the product of incoming messages.

$$\mu_{B \to f_2} = \mu_{f_0 \to B} * \mu_{f_3 \to B}$$

- $\mu_{f_3 \to B}$ is assumed one.
- $\mu_{f_0 \to B}$ is f_0(B)
- Which makes $\mu_{B \to f_2} = f_0$(B)

Now let's write messages for F.

From f_1 and f_4 there are incoming messages to F

From F to f_2 there is an outgoing message which reaches to G

The outgoing message could be written as the product of incoming messages.

$$\mu_{F \to f_2} = \mu_{f_1 \to F} * \mu_{f_4 \to F}$$

- $\mu_{f_4 \to F}$ is assumed one.
- $\mu_{f_1 \to F}$ is f_1(F)
- Which makes $\mu_{F \to f_2} = f_1$(F)

The belief propagation message from B, F to G can be written as:

$$\mu_{B,F \to G} = f_2(\mathrm{G}, \mathrm{B}, \mathrm{F}) * \mu_{B \to f_2} * \mu_{F \to f_2}$$

$$= f_2(\mathrm{G}, \mathrm{B}, \mathrm{F}) * f_0(B) * f_1(F)$$