Machine Learning

Homework Probabilistic Graphical Models

Hamit Kavas

2019

Table 8.2 The joint distribution over three binary variables.

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

Marginal probability is for for the multiple variables were laid out together in a table then the marginal probability of one variable (*X*) would be the sum of probabilities for the other variable on the margin of the table.

Since marginal probability is:

$$P(x = a_i) = \sum_{y \in A_y} P(x = a_i, y)$$

We need to solve this equation for each *a* and *b* values.

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

P(b, c) = P(a=0, b, c) + P(a=1, b, c) = $\begin{cases} 0.192 + 0.192 = 0.384, when\ b = 0,\ c = 0 \\ 0.144 + 0.064 = 0.208, when\ b = 0,\ c = 1 \\ 0.048 + 0.048 = 0{,}096, when\ b = 1,\ c = 0 \\ 0.216 + 0.096 = 0.312, when\ b = 1,\ c = 1 \end{cases}$

P(a, c) = P(a, b=0, c) + P(a, b=1, c) = $\begin{cases} 0.192 + 0.048 = 0.240, when\ a = 0,\ c = 0 \\ 0.144 + 0.216 = 0.360, when\ a = 0,\ c = 1 \\ 0.192 + 0.048 = 0{,}240, when\ a = 1,\ c = 0 \\ 0.064 + 0.096 = 0.160, when\ a = 1,\ c = 1 \end{cases}$

Single probabilities for *a* and *b*:

P(a) = P(a, b=0) + P(a, b=1) = $\begin{cases} 0.6, when\ a = 0, \\ 0.4, when\ a = 1, \end{cases}$

$$P(b) = P(a=0, b) + P(a=0, b) = \begin{cases} 0.592, when\ b = 0, \\ 0.408, when\ b = 1, \end{cases}$$

What we know is multiplication of individual probabilities of *a* and *b* is not equal to their joint probability.

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

We can proof this as shown below.

$$[0.336 \ = P(a=0, b=0)] \neq [P(a=0)P(b = 0) = 0.6 * 0.592]$$

In order to prove conditional dependency, we firstly need to calculate probability of *c*.

$$P(c) = \textstyle\sum_{a,b=0,1} P(a, b, c) = \begin{cases} 0.480, when\ c = 0, \\ 0.520, when\ c = 1, \end{cases}$$

When *c* = 0;

$$P(a, b \mid c = 0) = \frac{P(a, b, c = 0)}{P(c = 0)}$$

If we solve this equation for each value of *a* and *b*

$$\frac{P(a = 0, b = 0, c = 0)}{P(c = 0)} = {}^{0.192}/_{0.480} = 0.4$$

$$\frac{P(a = 1, b = 0, c = 0)}{P(c = 0)} = {}^{0.192}/_{0.480} = 0.4$$

$$\frac{P(a = 0, b = 1, c = 0)}{P(c = 0)} = {}^{0.048}/_{0.480} = 0.1$$

$$\frac{P(a = 1, b = 1, c = 0)}{P(c = 0)} = {}^{0.048}/_{0.480} = 0.1$$

When *c* = 1;

$$P(a, b \mid c = 1) = \frac{P(a, b, c = 1)}{P(c = 1)}$$

If we solve this equation for each value of *a* and *b*

$$\frac{P(a = 0, b = 0, c = 1)}{P(c = 1)} = {}^{0.144}/_{0.520} = 0.276$$

$$\frac{P(a = 0, b = 1, c = 1)}{P(c = 1)} = {}^{0.216}/_{0.520} = 0.415$$

$$\frac{P(a = 1, b = 0, c = 1)}{P(c = 1)} = {}^{0.064}/_{0.520} = 0.123$$

$$\frac{P(a = 1, b = 1, c = 1)}{P(c = 1)} = {}^{0.096}/_{0.520} = 0.184$$

Similarly, we also have:

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

This equation above shows P(a, c) = P(a, b=0, c) + P(a, b=1, c). Similarly, we can obtain:

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

Hereby, we can easily verify the statement that P(a,b|c) = P(a|c)P(b|c):

$$[0.1 = P(a = 1, b = 1|c = 0)] = [P(a = 1|c = 0)P(b = 1|c = 0) = 0.5 * 0.2 = 0.1]$$

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.

In previous example we have already calculated P(a), P(a, c) and P(b | c);

P(a) = P(a, b=0) + P(a, b=1) = $\begin{cases} 0.6, when\ a = 0, \\ 0.4, when\ a = 1, \end{cases}$

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

We can also obtain P(c | a):

$$P(c \mid a) = \frac{P(a,c)}{P(a)} = \begin{cases} 0.24/_{0.60} = 0.4, if\ a = 0, c = 0 \\ 0.36/_{0.60} = 0.6, if\ a = 0, c = 1 \\ 0.24/_{0.40} = 0.6, if\ a = 1, c = 0 \\ 0.16/_{0.40} = 0.4, if\ a = 1, c = 1 \end{cases}$$

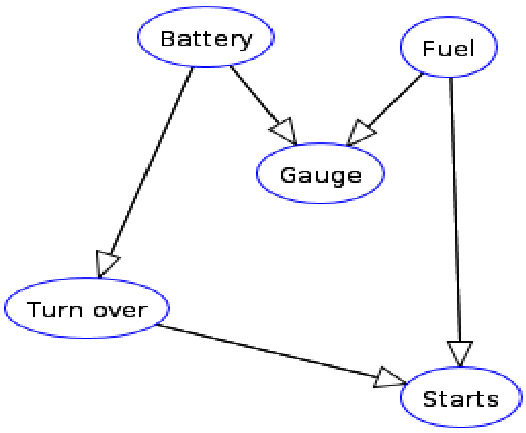
So it is clear that P(a, b, c) = P(a)P(c | a)P(b | c)

Directed graph is:

A → C → B

Answer-2

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



Answer a)

To begin with I will show general formula of conditional probabilities.

$$P(X_1, X_2, X_3, X_4, X_5) = \prod_{i=1}^{5} (X_i \mid X_1, X_2, \dots, X_{i-1})$$

We multiply the conditional probabilities by exploiting the conditional independence relationships in the graphical model

$$P(B, F, G, T, S) = P(B)P(F)P(G \mid B, F)\ P(T|B)P(S|F, T)$$

A joint distribution is a factor which returns a number which is the probability of a given combination of assignments.[1]

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

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

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

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

Answer b)

But with the graphical model, the computation can be much simpler.

i) $T \perp\!\!\!\perp F \,/\, \emptyset$?

What we need to do is analyze all possible paths between *Turn over* and *Fuel*.

First path would be $T \rightarrow S \leftarrow F$:

- $p(T, S, F) = p(T)p(F)p(S \mid T, F)$
- $\sum_S p(T, S, F) = \sum_S p(T)p(F)p(S \mid T, F)$
- $p(T, F) = p(T)p(F) \sum_S p(S \mid T, F)$
- $p(T, F) = p(T)p(F)$

Therefore, $\mathrm{T}, \mathrm{F} \mid \emptyset$ is blocked.

This proves that first path is blocked.

Second path would be $T \leftarrow B \rightarrow G \leftarrow F$. We will analyze this path in two steps.

Step 1 $F \rightarrow G \leftarrow B$:

- $p(F, G, B) = p(F)p(B)p(G \mid F, B)$
- $\sum_S p(F, G, B) = \sum_G p(F)p(B)p(G \mid F, B)$
- $p(F, B) = p(F)p(B) \sum_G p(G \mid F, B)$
- $p(T, F) = p(T)p(F)$

Therefore, $\mathrm{F{,}B} \mid \emptyset$ where it is blocked.

Step 2 $T \rightarrow G \leftarrow B$:

- $p(T, B, G) = p(T \mid B)p(G \mid B)p(B)$
- $\frac{p(T,B,G)}{p(B)} = \frac{p(T \mid B)p(G \mid B)p(B)}{p(B)}$
- $p(T, B, G) = p(T \mid B)p(G \mid B)$

Therefore, $\mathrm{T{,}G} \mid \mathrm{B}$ is non $-$ blocked.

$F \rightarrow G \leftarrow B$ on the second path is blocked, even though we have a possible pass through $T \rightarrow G \leftarrow B$, we must consider this path as also blocked.

Considering all possible paths are blocked we can say that *T* independent of *F* without any evidence.

The answer is Yes

<u>ii) T ⫫ F | S?</u>

Here we can discover this relation by looking one single path between them.

$T \rightarrow S \leftarrow F$:

- $p(T, S, F) = p(T)p(F)p(S \mid T, F)$
- $\frac{p(T,S,F)=}{p(S)} = \frac{p(T)p(F)p(S \mid T,F)}{p(S)}$
- $p(T, S, F) \neq p(T \mid S)p(F \mid S)$

Therefore, $\mathrm{T{,}F} \mid \mathrm{S}$ is non $-$ blocked.

So that *T* and *F* for given *S,* is not independent.

The answer is No

<u>iii) B ⫫ F | S?</u>

Since we already now that B→ $G \leftarrow F$ is blocked, we will only analyze path B→ $T \rightarrow S \leftarrow F$.

Since we also know that $T \rightarrow S \leftarrow F$ is non-blocked, we just need to analyze B→ $T \rightarrow S$:

- $p(B, T, S) = p(B)p(T \mid B)p(S \mid T)$

- $\frac{p(B,T,S)=}{p(T)} = \frac{p(B)p(T \mid B)p(S \mid T)}{p(T)}$
- $p(B,S \mid T) = p(B \mid T)p(S \mid T)$

Therefore, B,S | T is non − blocked.

Hereby, we can conclude that B and F for given S is non-independent.

The answer is No

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

$$p(B = bad) = 0.02 \qquad p(F = empty) = 0.05$$
$$p(G = empty|B = good, F = notempty) = 0.04 \qquad p(G = empty|B = good, F = empty) = 0.97$$
$$p(G = empty|B = bad, F = notempty) = 0.1 \qquad p(G = empty|B = bad, F = empty) = 0.99$$
$$p(T = false|B = good) = 0.03 \qquad p(T = false|B = bad) = 0.98$$
$$p(S = no|T = true, F = notempty) = 0.01 \qquad p(S = no|T = true, F = empty) = 0.92$$
$$p(S = no|T = false, F = notempty) = 1 \qquad p(S = no|T = false, F = empty) = 0.99$$

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). **You are encouraged to download and use the Java tool http://aispace.org/bayes/ to check numerically that your results are correct.**

Answer c)

Firstly, I will show the factors and their probabilistic definitions:

$$f_0 = P(Battery)$$

$$f_1 = P(Fuel)$$

$$f_2 = P(Gauge \mid Battery, Fuel)$$

$$f_3 = P(Turn\ over \mid Battery)$$

$$f_4 = P(Starts \mid Fuel, Turn\ over)$$

$$f_6 = P(Turn\ over)$$

$$f_7 = P(Fuel = F)$$

$$f_8 = P(Fuel = F,\ \ Starts = F)$$

Here is a brief explanation which tells True-False decisions for each individual node.

Nod/Boolean	*False*	True
Battery	*bad*	good
Fuel	*empty*	notempty
Gauge	*empty*	notempty
Turn Over	*True*	False
Starts	*Yes*	No

What is being asked to result of f_8.

$$\sum_{B,F,G,T,S} p(B,F,G,T,S)$$

To begin with, we will use our initial observation which is Starts = no.

$$\sum_{B,F,G,T} p(B,F,G,T,S=no)$$

And also Fuel = empty

$$\sum_{B,G,T} p(B,F=empty,G,T,S=no)$$

We can split summations to simplify the equation;

$$\sum_{B} p(B)p(F=empty)\sum_{G} p(G|B,F=empty)\sum_{T} p(T|B)p(S=no|F=empty,)$$

$$=\sum_{B\{bad,good\}} p(B)p(F=empty)\sum_{G\{empty,notempty\}} p(G|B,F=empty)\ldots$$

$$\ldots\sum_{T\{true,false\}} p(T|B)p(S=no|F=empty,)$$

In following equations, I will use B for B{good} and $\neg B$ for B{bad} and same way of showing for all other probabilities.

$$=p(\neg B)p(\neg F)[\,p(\neg G|\neg F,\neg B)[\,p(T|\neg B)p(\neg S|\neg F,T)+p(\neg T|\neg B)p(\neg S|\neg F,\neg T)]..$$

$$..+p(G|\neg F,\neg B)[\,p(T|\neg B)p(\neg S|\neg F,T)+p(\neg T|\neg B)p(\neg S|\neg F,\neg T)]]+\ldots$$

$$\ldots p(B)p(\neg F)[[\,p(\neg G|\neg F,B)[\,p(T|B)p(\neg S|\neg F,T)+p(\neg T|B)p(\neg S|\neg F,\neg T)]\ldots$$

$$\ldots+p(G|\neg F,B)[\,p(T|B)p(\neg S|\neg F,T)+p(\neg T|B)p(\neg S|\neg F,\neg T)]]$$

When we add the numerical values, we get;

$$=0.02*0.05[0.99*[(0.02*0.92)+(0.98*0.99)]+0.01*[(0.92*0.02)+(0.99*0.98)]$$

$$+0.98*0.05[0.97*(0.92*0.97)+(0.99*0.03)]+0.03*[(0.92*0.97)+(0.99*0.03)]=0.0461715$$

As next step, we can simplify our equation by eliminating the tail probabilities that equals to one. In our example we have

$$p(\neg G|\neg F,\neg B)=0.99$$

$$p(G|\neg F,\neg B)=0.01$$

And;

$$p(\neg G|\neg F, B) = 0.97$$

$$p(G|\neg F, B) = 0.03$$

Considering we sum product with these two couples with same values, we can subtract them from our equation.

$$p(\neg B)p(\neg F)\ [p(T|\neg B)p(\neg S|\neg F, T) + p(\neg T|\neg B)p(\neg S|\neg F, \neg T)] \ldots$$

$$\ldots + \ p(B)p(\neg F)[p(T|B)p(\neg S|\neg F, T) + p(\neg T|B)p(\neg S|\neg F, \neg T)]$$

With numerical values;

$$0.02 * 0.05(0.02 * 0.92 + 0.98 * 0.99) + 0.98 * 0.05(0.97 * 0.92 + 0.03 * 0.99)$$

$$= 0.0461715$$

We got the same result with 6 terms.

$$p(S = no) = \sum_{F} p(F, S = no)$$

$$= p(F = empty, S = no) + \ p(F = notempty, S = no)$$

$$= 0.0461715 + p(F = notempty, S = no)$$

What we need to do is just to calculate the same equation for *Fuel = notempty:*

$$\sum_{B} p(B)p(F = notempty) \sum_{G} p(G|B, F = notempty) \ldots$$

$$\ldots \sum_{T} p(T|B)p(S = no|F = notempty)$$

$$= \sum_{B\{bad,good\}} p(B)p(F = notempty) \sum_{G\{empty,notempty\}} p(G|B, F = notempty) \ldots$$

$$\ldots \sum_{T\{true,false\}} p(T|B)p(S = no|F = notempty)$$

$$p(\neg B)p(F)\ [p(T|\neg B)p(\neg S|F, T) + p(\neg T|\neg B)p(\neg S|F, \neg T)] \ldots$$

$$\ldots + \ p(B)p(F)[p(T|B)p(\neg S|F, T) + p(\neg T|B)p(\neg S|F, \neg T)]$$

With numerical values;

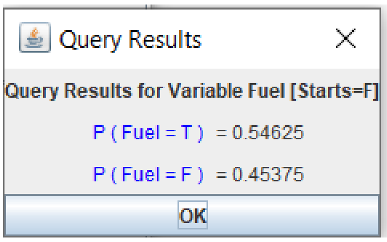
$$0.02 * 0.95 * [(0.02 * 0.01) + (0.98 * 1)] + 0.98 * 0.95 * [(0.97 * 0.01) + 0.03 * 1)]$$

= 0.0555845

Finally, by marginalizing the asked equation we would find the normalized result.

$$p(F = empty \mid S = no) = \frac{p(F = empty, S = no)}{p(S = no)} = \frac{0.0461715}{0.0461715 + 0.0555845} = 0.4537$$

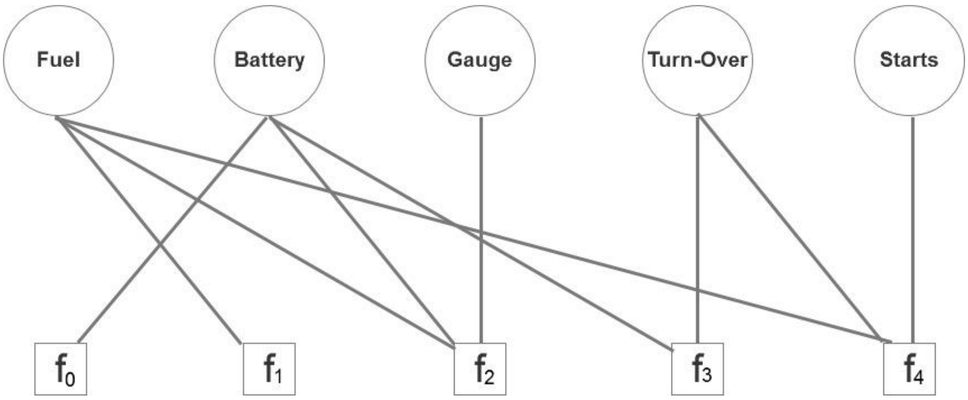
We confirm that our results are correct by using Java tool:



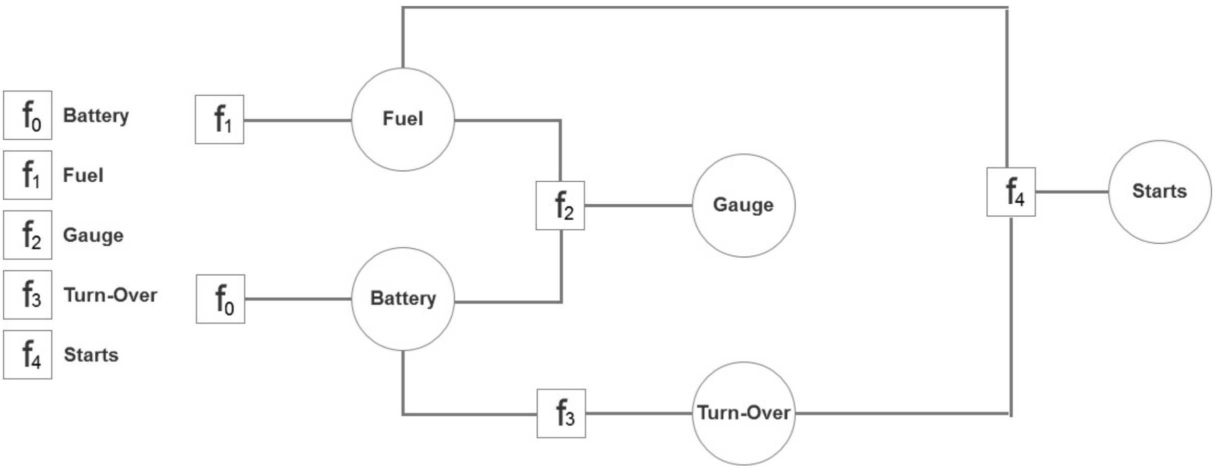
(d) Transform the model into an equivalent factor graph and describe it fully, e.g., write down all the involved factors and their values.

Answer d)

Factor is a function which takes all possible combinations of outcomes by including the appropriate features.[3] The factor graph associated with the function[4]:



We select factors according to relations between nodes. I have named all factors with the node names due to every node is being created by one individual factor. Factor graph is below:



Each factor function has some values that associated with it, which describes how much influence the factor has on its variables in relative terms. In other words, the values encode the confidence we have in the relationship expressed by the factor function.[5] In order to calculate values, probability values must be initiated.

B	f_0 (B)
0	0.02
1	0.98

F	f_1 (F)
0	0.05
1	0.95

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

T	B	f_3 (T,B)
0	1	0.03
1	1	0.97
0	0	0.98
0	0	0.02

S	T	F	f_4(F,B,G)
0	1	0	0.01
1	1	0	0.99
0	1	1	0.92
0	1	1	0.08
1	0	0	1.00
1	0	0	0.00
0	0	1	0.99
1	0	1	0.01

(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 e)

We need to find messages that are carried by each factor that affects B,F and G.

Initial message from f_1 to Fuel:

$$ɥ f_1 \to F = f_1$$

Message coming back from f_4 to Fuel:

$$ɥ f_4 \to F = 1$$

Message going from Fuel to f_2 equals to multiplication of other factors that affect *F*:

$$ɥ B \to f_2 = (ɥ f_1 \to F) * (ɥ f_4 \to F) = f_1$$

Initial message from f_0 to Battery:

$$\mu f_0 \to B = f_0$$

Message coming back from f_3 to Battery:

$$\mu f_3 \to F = 1$$

Message going from Battery to f_2 equals to multiplication of other factors that affect B:

$$\mu B \to f_2 = (\mu f_0 \to F) * (\mu f_3 \to F) = f_0$$

Under this circumstances, belief propagation messages from B, F to G are shown below.

$$\mu B, F \to G = f_2(G, B, F) * (\mu B \to f_2) * (\mu F \to f_2)$$

$$= f_2(G, B, F) * f_0(B) * f_1(F)$$

We have calculated beliefs in Python as shown below.

Fuel belief (Fuel=True) : [0.0016222 0.9983778]

Battery belief (Battery=True): [0.01875397 0.98124603]

Fuel belief (Fuel=False) : [0.55350217 0.44649783]

Battery belief (Battery=False): [0.03296829 0.96703171]

References

1-https://frnsys.com/ai_notes/foundations/probabilistic_graphical_models.html

2-https://www.cs.princeton.edu/courses/archive/spring07/cos424/scribe_notes/0308.pdf

3- https://frnsys.com/ai_notes/foundations/probabilistic_graphical_models.html

4-d mcay, information theory

5- http://deepdive.stanford.edu/inference