COMP7703 Machine Learning
Assignment 4

Thanat Chokwijitkul 44522328

Question 10.3

Implement the Bayesian network from the "Research Assistance" example in the notes. Use the probability values given and verify the result.

Suppose it is Saturday and a research assistance is working in his windowless office as usual. He is looking forward to a baseball game later, but he has some concern about the likelihood of rain, which could lead to the game being cancelled. In addition to that, one of the research assistance's kids is at the beach, and he is wondering whether the child might get sunburn.

Let C, R, S, G denote clouds, rain, sunburn and the baseball game respectively. Table 1 contains four tables outlining the prior probabilities of clouds ($P(C)$) along with the conditional probabilities of rain and the research assistance's kid getting sunburn given the presence and the absence of clouds ($P(R|C)$ and $P(S|C)$), and the game being played given the presence and the absence of rain ($P(G|R)$).

Table 1: The probability values obtained from Figure 10.4 in the lecture note

Clouds	
$P(C)$	0.1
$P(\neg C)$	0.9

Rain	
$P(R\|C)$	0.6
$P(\neg R\|C)$	0.4
$P(R\|\neg C)$	0.0
$P(\neg R\|\neg C)$	1.0

Sunburn	
$P(S\|C)$	0.1
$P(\neg S\|C)$	0.9
$P(S\|\neg C)$	0.7
$P(\neg S\|\neg C)$	0.3

PlayGame	
$P(G\|R)$	0.05
$P(\neg G\|R)$	0.95
$P(G\|\neg R)$	1.0
$P(\neg G\|\neg R)$	0.0

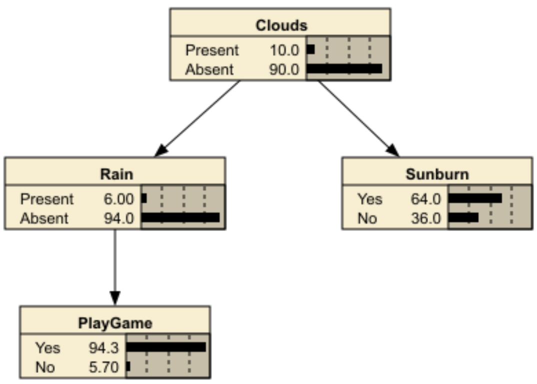

Figure 1: The Bayesian network representing the "research assistance" case scenario

Using the probability values in Table 1, the case scenario can be represented by the Bayesian belief network as shown in Figure 1. The root node is the prior probability of cloud in the area, which is expected to be cloudy 10% of the time. The Rain node contains the marginal probabilities for rain given the presence and the absence of cloud. Similarly, the Sunburn node contains the marginal probabilities for the research assistance's kid getting sunburn given the presence and the absence of cloud. Lastly, the PlayGame node contains the marginal probabilities for the upcoming baseball game whether it is going to be played or not. The probabilities presented in Figure 1 are identical to the marginal probabilities presented in Figure 10.4 in the lecture note.

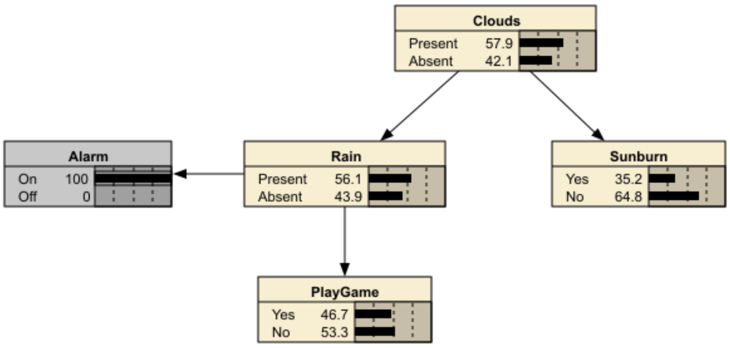

Figure 2: The Bayesian network representing the "research assistance" case scenario assuming the rain alarm goes off

Table 2: The probability values if the rain alarm goes off

Clouds		Rain		Sunburn		GamePlay	
Present	0.58	Present	0.56	Yes	0.35	Yes	0.47
Absent	0.42	Absent	0.44	No	0.65	No	0.53

Suppose that the research assistance's office has a rain alarm installed and it goes off in the afternoon. Let A denotes the rain alarm. According to the lecture note, $P(A|R) = 0.8$ and $P(A|\neg R) = 0.04$. After adding an extra node for the rain alarm and because it goes off in the afternoon, the probabilities of each node can be updated as shown in Figure 2 and Table 2, which are identical to the probabilities presented in Figure 10.5 in the lecture note.

When the alarm goes off, the probability of the presence of rain increases to 0.56, which also infers to a higher probability of the area being cloudy (0.58). An increased amount of clouds results in a lower chance of the research assistance's kid getting sunburn (0.35).

(a) If the rain alarm does NOT go off, what are the probabilities for the other nodes? Does this make sense to you? Explain in a couple of sentences.

Table 3: The probability values if the rain alarm does not goes off

Clouds		Rain		Sunburn		GamePlay	
Present	0.05	Present	0.01	Yes	0.67	Yes	0.99
Absent	0.95	Absent	0.99	No	0.33	No	0.01

As opposed to the scenario illustrated by Figure 2, Figure 3 is the Bayesian network representing the situation when there is a rain alarm installed in the research assistance's windowless office and it does NOT go off, and Table 3 shows the updated probability values.

It is clear that the rain alarm node has had the effect of changing the probabilities in other nodes. The updated set of probability values is considered reasonable since when the alarm does not go off, the probability of the presence of rain decreases to 0.01, which also infers to a small chance of the area being cloudy (0.05). A decreased amount of clouds results in a higher chance of the research assistance's kid getting sunburn (0.67).

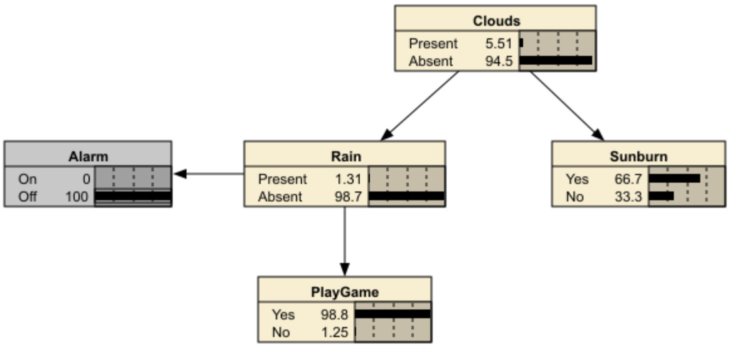

Figure 3: The Bayesian network representing the "research assistance" case scenario assuming the rain alarm does not goes off

Question 10.4

From slide 10 in the lecture notes, calculate $P(C|W)$;

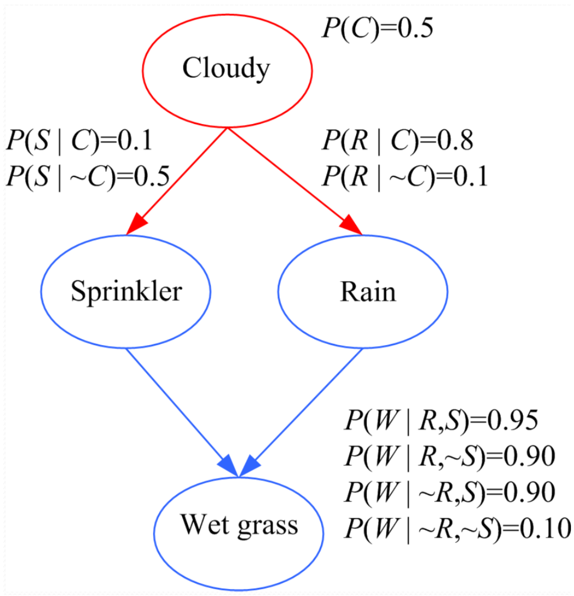

Figure 4: The Bayesian network from slide 10 in the lecture notes

(a) Work the answer out by hand (using pen and paper, showing your working).

Let C, S, R, W denote cloudy, sprinkler, rain and wet grass respectively. The joint probability function of the Bayesian belief network depicted in Figure 4 is:

$$P(C, S, R, W) = P(C)P(S|C)P(R|C)P(W|S, R)$$

In order to compute the probability of the sky being cloudy given that the grass is wet, we can apply Bayes' rule as follow:

$$P(C|W) = \frac{P(W|C)P(C)}{P(W)}$$

The probability of the grass being wet given the presence of cloud, denoted as $P(W|C)$, can be computed as follows:

$$\begin{aligned}
P(W|C) &= P(W|R,S)P(R,S|C) + P(W|\neg R,S)P(\neg R,S|C) \\
&\quad + P(W|R,\neg S)P(R,\neg S|C) + P(W|\neg R,\neg S)P(\neg R,\neg S|C) \\
&= P(W|R,S)P(R|C)P(S|C) + P(W|\neg R,S)P(\neg R|C)P(S|C) \\
&\quad + P(W|R,\neg S)P(R|C)P(\neg S|C) + P(W|\neg R,\neg S)P(\neg R|C)P(\neg S|C) \\
&= 0.95 \cdot 0.8 \cdot 0.1 + 0.9 \cdot 0.2 \cdot 0.1 + 0.9 \cdot 0.8 \cdot 0.9 + 0.1 \cdot 0.2 \cdot 0.9 \\
&= 0.76
\end{aligned}$$

The probability of the grass being wet, denoted as $P(W)$, can be computed as follows:

$$\begin{aligned}
P(W) &= \sum_{C,R,S \in \{T,F\}} P(C,R,S,W) \\
&= \sum_{C,R,S \in \{T,F\}} P(C)P(R|C)P(S|C)P(W|R,S) \\
&= P(C)P(R|C)P(S|C)P(W|R,S) + P(C)P(\neg R|C)P(S|C)P(W|\neg R,S) \\
&\quad + P(C)P(R|C)P(\neg S|C)P(W|R,\neg S) + P(C)P(\neg R|C)P(\neg S|C)P(W|\neg R,\neg S) \\
&\quad + P(\neg C)P(R|\neg C)P(S|\neg C)P(W|R,S) + P(\neg C)P(\neg R|\neg C)P(S|\neg C)P(W|\neg R,S) \\
&\quad + P(\neg C)P(R|\neg C)P(\neg S|\neg C)P(W|R,\neg S) + P(\neg C)P(\neg R|\neg C)P(\neg S|\neg C)P(W|\neg R,\neg S) \\
&= 0.5 \cdot 0.8 \cdot 0.1 \cdot 0.95 + 0.5 \cdot 0.2 \cdot 0.1 \cdot 0.9 + 0.5 \cdot 0.8 \cdot 0.9 \cdot 0.9 + 0.5 \cdot 0.2 \cdot 0.9 \cdot 0.1 \\
&\quad + 0.5 \cdot 0.1 \cdot 0.5 \cdot 0.95 + 0.5 \cdot 0.9 \cdot 0.5 \cdot 0.9 + 0.5 \cdot 0.1 \cdot 0.5 \cdot 0.9 + 0.5 \cdot 0.9 \cdot 0.5 \cdot 0.1 \\
&= 0.65125
\end{aligned}$$

Having known $P(C)$, $P(W|C)$ and $P(W)$, the probability of the sky being cloudy given the grass is wet, denoted as $P(C|W)$, can be computed as follows:

$$\begin{aligned}
P(C|W) &= \frac{P(W|C)P(C)}{P(W)} \\
&= \frac{0.76 \cdot 0.5}{0.65125} \\
&= 0.5834932821497121 \\
&\approx 0.584
\end{aligned}$$

Thus, the probability of the sky being cloudy given the grass is wet is approximately 0.584.

(b) Use Netica to simulate this network and calculate the answer.

Note: In addition to simulating the network and calculating the answer for $P(C|W)$, Netica was also used to verify the values of $P(W|C)$ and P(W).

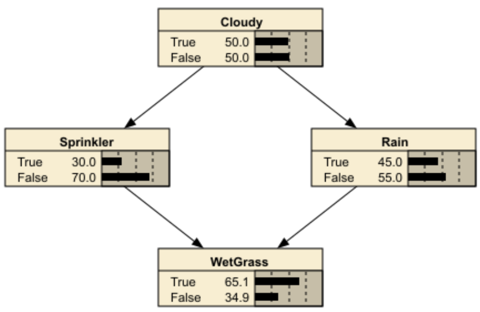

Figure 5: The Bayesian network from slide 10 in the lecture notes

Figure 5 is the Bayesian network from slide 10 in the lecture notes. The prior probabilities for the Cloudy node and the conditional probabilities for the other nodes have been assigned to the network based on the probability values in Figure 4. According to Figure 5, the probability of the grass being wet, namely $P(W)$ is 0.65, which is equal to the $P(W)$ value calculated in Question 10.4(a).

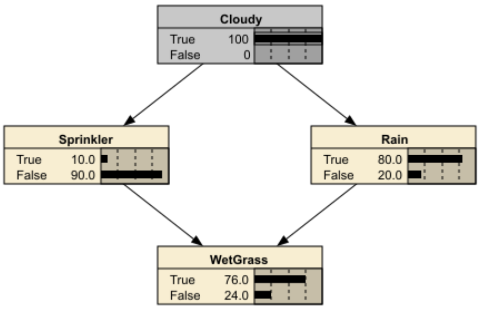

Figure 6: The Bayesian network from slide 10 in the lecture notes assuming that the sky is cloudy

The Bayesian network in Figure 6 assumes that the sky is cloudy. From Figure 6, it can be observed that the probability of the grass being wet given the sky is cloudy, namely $P(W|C)$ is 0.76, which is equal to the $P(W|C)$ value calculated in Question 10.4(a).

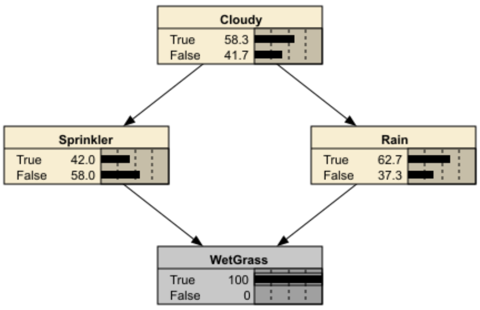

Figure 7: The Bayesian network from slide 10 in the lecture notes assuming that the grass is wet

Ultimately, Figure 7 is what the question asks for. It simulates the network assuming that the grass is already wet. According to Figure 7, the probability of the sky being cloudy given the grass is wet, namely $P(C|W)$, is 0.58, which is also equal to the $P(C|W)$ value calculated in Question 10.4(a).

It can be observed that if the grass is wet, the chance of rain is likely to become higher along with the probability of the sprinkler being on, which infers to higher probability of the sky being cloudy.

Question 11.1

Implement Bayesian learning for the case where the data is 1D (i.e. a single continuous variable) and the data is assumed to follow a Gaussian distribution with known/constant variance (i.e. the example in Alpaydin, Section 16.3.1 and discussed in lectures). Specifically, your code should use the third column from the Iris dataset and produce a graph of the model prior and posterior distributions. Add the data as well: this is essentially reproducing Fig.16.3 from Alpaydin, plotting the data instead of $p(x|\mu)$.

```
load('iris.txt')
X = iris(:,3);
N = length(X);

mu_0 = 2;
sigma_0 = 0.5;

m = mean(X);
sigma = std(X);

mu_N = ((sigma^2/(N*sigma_0^2+sigma^2))*mu_0)+((N*sigma_0^2/(N*sigma_0^2+sigma^2))*m);
sigma_N_squared = 1/(1/sigma_0^2+N/sigma^2);
sigma_N = sqrt(sigma_N_squared);

x = 0:0.01:7;
prior_dist = normpdf(x,mu_0,sigma_0);
posterior_dist = normpdf(x,mu_N,sigma_N);

hold all
plot(x,prior_dist)
plot(x,posterior_dist)
plot(X,0,'*k')
title 'Prior and Posterior Distributions'
legend('Prior','Posterior','Data point')
xlabel('x')
ylabel('y')
```

The MATLAB code above implements Bayesian parameter estimation for the case of one-dimensional data that follows a Gaussian distribution with unknown mean and known variance.

Note: The `normpdf` function takes a standard deviation as its input.

Precisely, the code above uses the petal length data (the third column) from the Iris dataset, denoted by $\mathcal{X} = \{x^t\}_{t=1}^N$ where N is the number of samples and x^t are continuous variables drawn from a Gaussian distribution with known variance. This univariate case can be written as $p(x) \sim \mathcal{N}(\mu, \sigma^2)$, where μ and σ^2 are parameters.

The prior for μ is also a Gaussian distribution, denoted by $p(\mu) \sim \mathcal{N}(\mu_0, \sigma_0^2)$, where $p(\mu)$ is the prior density, μ_0 is the prior mean and σ_0^2 is the prior variance. Instead of using an uninformative prior [1], assume we know that the averaged petal length of most of the Iris species is 2 cm with a variance of 0.25 cm (a standard deviation of 0.5), we have $\mu_0 = 2$, $\sigma_0^2 = 0.25$ ($\sigma_0 = 0.5$).

According to [2], using the prior information, the posterior can be written as $p(\mu|\mathcal{X}) \sim \mathcal{N}(\mu_N, \sigma_N^2)$ which is still a Gaussian where

$$\mu_N = \frac{\sigma^2}{N\sigma_0^2 + \sigma^2}\mu_0 + \frac{N\sigma_0^2}{N\sigma_0^2 + \sigma^2}m$$

$$\frac{1}{\sigma_N^2} = \frac{1}{\sigma_0^2} + \frac{N}{\sigma^2}$$

where $m = \frac{1}{N}\sum_{t=1}^{N} x^t$ is the sample average. The implementation of the formulas for μ_N and σ_N can be found in line 10 and 11 respectively.It can be observed that the mean μ_N is a weighted average of μ_0 and m, with weights being inversely proportional to their variances. It can be observed that the mean μ_N is a weighted average of μ_0 and m, with weights being inversely proportional to their variances. The prior parameters (μ_0 and σ_0) and the posterior parameters (μ_N and σ_N) will subsequently be used to compute the probability density function (PDF) of both prior and posterior distributions, which can be plotted as shown in Figure 8.

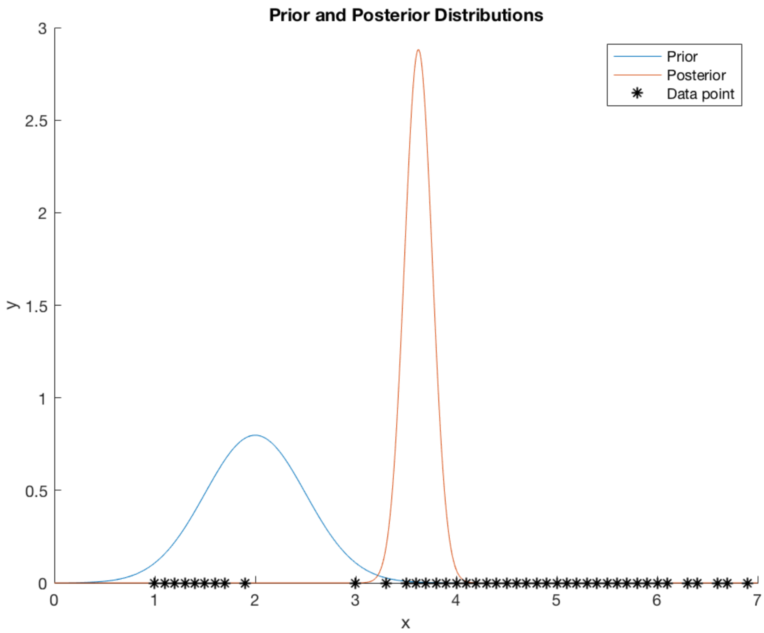

Figure 8: Prior and posterior distributions

Figure 8 is a plot produced by the above MATLAB script illustrating the prior and posterior distributions along with the data points obtained from the third column of the Iris dataset. One hundred and fifty data points are drawn from $p(x) \sim \mathcal{N}(3.76, 1.76^2)$, prior is $p(\mu) \sim \mathcal{N}(2, 0.5^2)$ and posterior is $p(\mu|\mathcal{X}) \sim \mathcal{N}(3.62, 0.14^2)$. It can be observed that the mean of the posterior density μ_N is larger than the prior mean μ_0 and close to the sample mean m. However, the posterior variance σ_N is smaller than both the prior variance σ_0 and $\sigma/\sqrt{N}$, which is consistent with the statement "σ_N gets smaller when either of σ_0 or σ gets smaller or if N is larger" mentioned in [2], Section 16.3.1.

References

[1] R. Yang and J. O. Berger, *A catalog of noninformative priors.* Institute of Statistics and Decision Sciences, Duke University, 1996.

[2] E. Alpaydin, *Introduction to machine learning.* MIT press, 2014.

[3] S. Marsland, *Machine learning: an algorithmic perspective.* CRC press, 2015.

[4] K. P. Murphy, *Machine learning: a probabilistic perspective.* MIT press, 2012.

[5] T. Downs, "Lecture 10 - Bayesian Belief Networks," School of Information Technology and Electrical Engineering (ITEE), The University of Queensland.