1. Given the following Bayesian network of Boolean random variables.
   Let $P(I)$=0.1, $P(D)$=0.2, $P(S|I) = 0.3$, $P(S|\neg I) = 0.4$,
   $P\ (G|I, D) = 0.1$, $P\ (G|I, \neg D) = 0.2$, $P(G|\neg I,D)$=0.3,$P(G|\neg I,\neg D)$=0.4,
   $P(R|G) = 0.1$, $P(R|\neg G) = 0.2$.
   a. Use the Variable Elimination algorithm to compute $P(R)$. Eliminate variables in the order of D, S, I, and G. Show intermediate steps.

$$P(R) = \sum_G \sum_I \sum_S \sum_D P(D)P(I)P(S|I)P(G|I,D)P(R|G)$$
$$P(R) = \sum_G P(R|G) \sum_I P(I) \sum_S P(S|I) \sum_D P(D)P(G|I,D)$$

$P(I)$	$P(D)$	$P(G)$	$P(\neg G)$
T	T	0.1	0.9
T	F	0.2	0.8
F	T	0.3	0.7
F	F	0.4	0.6

D	$P(D)$
T	0.2
F	0.8

$P(D)P(G|I,D) = F(G,I,D) =$

$P(I)$	$P(D)$	$F(G,I,D)$	$F(\neg G,I,D)$
T	T	0.02	0.18
T	F	0.16	0.64
F	T	0.06	0.14
F	F	0.32	0.48

$$P(R) = \sum_G P(R|G) \sum_I P(I)F(G,I) \sum_S P(S|I)$$

where $F(G,I) =$

$P(I)$	$F(G,I,D)$	$F(\neg G,I,D)$
T	0.18	0.82
F	0.38	0.62

Is equivalent to

$P(I)$	$P(G)$	$P(G|I)$
T	T	0.18
T	F	0.82
F	T	0.38
F	F	0.62

$$\sum_S P(S|I) = 1$$

$$\therefore P(R) = \sum_G P(R|G) \sum_I P(I) F(G,I)$$

I	$P(I)$
T	0.1
F	0.9

$$\sum_I P(I) F(G,I) = F(G,I) =$$

$P(I)$	$P(G)$	$P(G|I)$
T	T	0.018
T	F	0.082
F	T	0.342
F	F	0.558

$$P(R) = \sum_G P(R|G) F(G)$$

where $F(G) =$

G	$F(G)$
T	0.360
F	0.640

$P(R|G) =$

$P(R)$	$P(G)$	$P(R|G)$
T	T	0.1
T	F	0.2
F	T	0.9
F	F	0.8

$$\sum_G P(R|G) F(G) = F(R) = P(R) =$$

$P(R)$	$P(G)$	$P(R|G)$
T	T	0.036
T	F	0.128
F	T	0.324
F	F	0.512

$P(R) =$

R	$P(R)$
T	0.164
F	0.836

2. Use the Variable Elimination algorithm to compute $P(G)$. Eliminate variables in the order of D, S, I, and R. Show intermediate steps.

$$P(G) = \sum_R \sum_I \sum_S \sum_D P(D)P(I)P(S|I)P(G|I,D)P(R|G)$$
$$P(G) = \sum_R P(R|G) \sum_I P(I) \sum_S P(S|I) \sum_D P(D)P(G|I,D)$$

$P(I)$	$P(D)$	$P(G)$	$P(\neg G)$
T	T	0.1	0.9
T	F	0.2	0.8
F	T	0.3	0.7
F	F	0.4	0.6

D	$P(D)$
T	0.2
F	0.8

$P(I)$	$P(D)$	$F(G,I,D)$	$F(\neg G,I,D)$
T	T	0.02	0.18
T	F	0.16	0.64
F	T	0.06	0.14
F	F	0.32	0.48

$$P(G) = \sum_R P(R|G) \sum_I P(I)F(G,I) \sum_S P(S|I)$$

where $F(G,I) =$

$P(I)$	$F(G,I,D)$	$F(\neg G,I,D)$
T	0.18	0.82
F	0.38	0.62

Is equivalent to

$P(I)$	$P(G)$	$P(G|I)$
T	T	0.18
T	F	0.82
F	T	0.38
F	F	0.62

$$\sum_{S} P(S|I) = 1$$

$$\therefore P(G) = \sum_{R} P(R|G) \sum_{I} P(I) F(G, I)$$

I	$P(I)$
T	0.1
F	0.9

$$\sum_{I} P(I) F(G, I) = F(G, I) =$$

$P(I)$	$P(G)$	$P(G\|I)$
T	T	0.018
T	F	0.082
F	T	0.342
F	F	0.558

$$P(G) = \sum_{R} P(R|G) F(G)$$

where $F(G) =$

G	$F(G)$
T	0.360
F	0.640

$P(R|G) =$

$P(R)$	$P(G)$	$P(R\|G)$
T	T	0.1
T	F	0.2
F	T	0.9
F	F	0.8

$$\sum_{R} P(R|G) F(G) = F(G) = P(G) =$$

$P(R)$	$P(G)$	$P(R\|G)$
T	T	0.036
T	F	0.128
F	T	0.324
F	F	0.512

$P(G) =$

R	$P(R)$
T	0.36
F	0.64

Use the Variable Elimination algorithm to compute $P\ (G|I,S)$. Eliminate variables in the order of D and R. Show intermediate steps.

Given I, G and S are independent. Therefore

$$P(G|I,S) = P(G|I) = \frac{P(G,I)}{P(I)}$$

$$P(G,I) = \sum_R \sum_S \sum_D P(D)P(I)P(S|I)P(G|I,D)P(R|G)$$

$$P(G,I) = P(I) \sum_R P(R|G) \sum_S P(S|I) \sum_D P(D)P(G|I,D)$$

$P(I)$	$P(D)$	$P(G)$	$P(\neg G)$
T	T	0.1	0.9
T	F	0.2	0.8
F	T	0.3	0.7
F	F	0.4	0.6

D	$P(D)$
T	0.2
F	0.8

$P(I)$	$P(D)$	$F(G,I,D)$	$F(\neg G,I,D)$
T	T	0.02	0.18
T	F	0.16	0.64
F	T	0.06	0.14
F	F	0.32	0.48

$$P(G,I) = P(I)F(G,I) \sum_R P(R|G) \sum_S P(S|I)$$

where $F(G,I) =$

$P(I)$	$F(G,I,D)$	$F(\neg G,I,D)$
T	0.18	0.82
F	0.38	0.62

Is equivalent to

$P(I)$	$P(G)$	$P(G\|I)$
T	T	0.18
T	F	0.82
F	T	0.38
F	F	0.62

$$\sum_{S} P(S|I) = 1$$

$$P(G, I) = P(I)F(G, I) \sum_{R} P(R|G)$$

$$\sum_{R} P(R|G) = 1$$

$$P(G, I) = P(I)F(G, I)$$

$$P(G|I, S) = P(G|I) = \frac{P(G, I)}{P(I)} = \frac{P(I)F(G, I)}{P(I)} = F(G, I)$$

$P(G|I, S) = P(G|I) =$

$P(I)$	$P(G)$	$P(G\|I)$
T	T	0.18
T	F	0.82
F	T	0.38
F	F	0.62

Use the Direct Sampling algorithm to generate one sample. During sampling, choose the more likely value for each variable.

Sample obtained through direct sampling is,

$$\neg D, \neg I, \neg G, \neg S, \neg R$$

Given evidence I and S, use the Likelihood Weighting algorithm to generate one sample and its corresponding weight. During sampling, choose the more likely value for each variable.

Assuming that I, S indicates the case both of them are being positive. The sample obtained through likelihood weighting is

$$\neg D, I, \neg G, S, \neg R$$

The corresponding weight is $0.1 \times 0.3 = 0.03$

Apply Gibbs Sampling algorithm. Starting from initial $(\neg D, \neg I, \neg G, \neg S, \neg R)$, resample variables in the order of D, I, G, S, and R, to generate five samples. During sampling, choose the more likely value for each variable

Initial sample $\neg D, \neg I, \neg G, \neg S, \neg R$

$$\neg D, \neg I, \neg G, \neg S, \neg R$$
$$\neg D, \neg I, \neg G, \neg S, \neg R$$
$$\neg D, \neg I, \neg G, \neg S, \neg R$$
$$\neg D, \neg I, \neg G, \neg S, \neg R$$
$$\neg D, \neg I, \neg G, \neg S, \neg R$$

2. (2 points) Given the following training dataset for junk email detector.

SPAM:
CLICK HERE TO WIN MONEY
GAMBLING CLICK HERE

HAM:
GO TO LAS VAGAS CLICK HERE
LAS VAGAS IS GAMBLING CITY
GAMBLING COSTS MONEY

Using the Bag of Words method, what is the size of vocabulary that contains all of the words in the messages?
The size of the vocabulary is 13.

What is the ML (maximum likelihood) solution for learning $P(SPAM)$?

In our observed data we have 2 spam email, 3 ham emails, the data can be represented as SSHHH. We want to find the prior probability that maximizes the likelihood of observed data.

Let $P(S) = \pi \Rightarrow P(H) = 1 - \pi$. It can be represented in equation form as

$$P(y_i) = \pi^{y_i}(1-\pi)^{1-y_i}$$

Assuming that n data samples are drawn from independent, identical distribution

$$P(Y) = \prod_{i=1}^{n} P(y_i) = \prod_{i=1}^{n} \pi_i^{y}(1-\pi)^{1-y_i}$$

Taking log on both sides

$$\log P(Y) = \sum_{i=1}^{n} y_i \log \pi + (1-y_i)\log(1-\pi)$$

We want to find the value of π that maximizes the above equation. In order to get that value of π,we differentiate the above equation with π and equate it to zero.

$$\sum_{i=1}^{n} \frac{y_i}{\pi} + \frac{(1-y_i)}{1-\pi} = 0$$

Our observed data is as follows SSHHH, it can be represented as 11000.

Therefore, the above equation becomes

$$\frac{2}{\pi} + \frac{3}{1-\pi} = 0 \Rightarrow \pi = \frac{2}{5}$$

which is the maximum likelihood estimate of spam.

Using Naïve Bayes model and ML learning, what is the ML solution for $P(\text{"CLICK"}|SPAM)$?

$$P(\text{CLICK}|SPAM) = \frac{2}{8} = 0.25$$

Using Naïve Bayes model and ML learning, for message M= "GAMBLING IS HERE", What is $P(SPAM|M)$?

$$P(SPAM|M) = \frac{P(M|SPAM)}{P(M)}$$
$$= \frac{P(\text{GAMBLING IS HERE}|SPAM) \times P(SPAM)}{P(\text{GAMBLING IS HERE}|SPAM)\times P(SPAM) + P(\text{GAMBLING IS HERE}|HAM)\times P(HAM)}$$

$$P(\text{GAMBLING IS HERE}|SPAM)\times P(SPAM) = \frac{1}{8}\times 0\times\frac{1}{4}\times\frac{2}{5} = 0$$
$$\therefore P(SPAM|M) = 0$$

If we use Laplace smoothing with k = 1

$$P(SPAM) = \frac{2+1}{5+2} = \frac{3}{7}, \qquad P(HAM) = \frac{4}{7}$$
$$P(\text{GAMBLING}|SPAM) = \frac{1+1}{8+13} = \frac{2}{21}$$
$$P(\text{IS}|SPAM) = \frac{1}{8+13} = \frac{1}{21}$$
$$P(\text{HERE}|SPAM) = \frac{2+1}{8+13} = \frac{3}{21}$$
$$P(\text{GAMBLING}|HAM) = \frac{2+1}{14+13} = \frac{3}{27}$$
$$P(\text{IS}|HAM) = \frac{1+1}{14+13} = \frac{2}{27}$$
$$P(\text{HERE}|HAM) = \frac{1+1}{14+13} = \frac{2}{27}$$

$$P(SPAM|M) = \frac{\frac{2}{21}\times\frac{1}{21}\times\frac{3}{21}\times\frac{3}{7}}{\frac{2}{21}\times\frac{1}{21}\times\frac{3}{21}\times\frac{3}{7} + \frac{3}{27}\times\frac{2}{27}\times\frac{2}{27}\times\frac{4}{7}} = 0.4435$$

Facial Emotion Detection Considering Partial Occlusion of Face Using Bayesian Network

The paper proposes a novel emotion detection system based on facial, which works even in the case partial occlusion, using Bayesian Network. The experimental results show that the recognition rates with the proposed method outperforms the conventional methods for the cases of images with occlusion, images with no occlusion. Unlike the conventional methods, the proposed method finds a good subset of facial features for emotion detection.

The proposed method learns the structure of Bayesnet in two phases, namely internal phase and external phase. In internal phase K2 algorithm is used to learn the structure among the facial features. Only a subset of features is selected by exploiting the causal relations among the features, and structure between emotions and the selected facial features is learned in the external phase.

The authors have performed emotion detection with non-occluded image and occluded images. In the case of non-occluded images, the proposed method outperforms conventional methods in identifying all the emotions such as angry, surprise, happy, disgust etc. as the conventional methods overfits the training data by considering the relationship between emotions and all facial features where as the proposed method only considers the relationships between emotions and a subset of facial features. It obtained high recognition rates for happiness and surprise but the recognition rates for angry, sadness and disgust were low, this is because of the similarity in facial expression for those emotions.

In the case of occluded images, the authors detected the emotions without filling the gaps, filled the gaps after detecting the emotions. The recognition rates of the method were higher than the conventional methods. One prime assumption with the occluded images that the methods makes is that two or more occlusions of facial parts would not occur at the same time.

The authors used Japanese Female Facial Expression (JAFFE) database to verify the effectiveness of the proposed method. JAFFE database constitutes of facial images acted by ten Japanese female subjects, expressing different emotions as well as neutral face. The database consists a total of 213 images out of which 183 images were of different emotions, 30 images were of neutral face. The images were normalized such that the distance between inner corners of one's eye is 30 pixels.

Overall the paper was very well written and proposes a novel idea.

I chose this paper as they are dealing with the occlusion case which most of the papers were not considering.