1.

Instructions: Students can use an electronic calculator and a *black* or *blue* pen for answering closed-book examination.

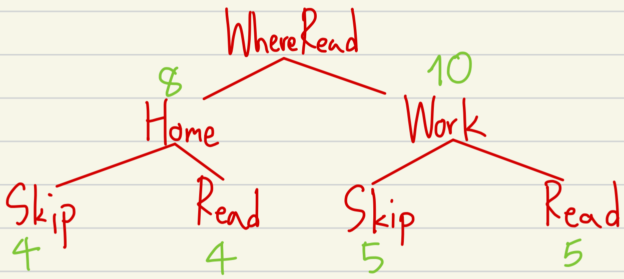
1. (4 marks) Consider the supervised learning dataset in **Figure 1**. Calculate the **log loss** of splitting the input attribute "***WhereRead***" in '***UserAction***' (Show all your calculation steps).

x split y

Example	*Author*	*Thread*	*Length*	*WhereRead*	*UserAction*
e_1	known	new	long	home	skips
e_2	unknown	new	short	work	reads
e_3	unknown	follow Up	long	work	skips
e_4	known	follow Up	long	home	skips
e_5	known	new	short	home	reads
e_6	known	follow Up	long	work	skips
e_7	unknown	follow Up	short	work	skips
e_8	unknown	new	short	work	reads
e_9	known	follow Up	long	home	skips
e_{10}	known	new	long	work	skips
e_{11}	unknown	follow Up	short	home	skips
e_{12}	known	new	long	work	skips
e_{13}	known	follow Up	short	home	reads
e_{14}	known	new	short	work	reads
e_{15}	known	new	short	home	reads
e_{16}	known	follow Up	short	work	reads
e_{17}	known	new	short	home	reads
e_{18}	unknown	new	short	work	reads
e_{19}	unknown	new	long	work	? Unseen
e_{20}	unknown	follow Up	long	home	?

Figure 1.

2. (2 marks) When the linear function has only **one** variable, show that its graph is a hyper-plane of



$$\text{log\_loss} = \frac{-\left(4\times\log_2\left(\frac{4}{8}\right) + 4\times\log_2\left(\frac{4}{8}\right) + 5\times\log_2\left(\frac{5}{10}\right) + 5\times\log_2\left(\frac{5}{10}\right)\right)}{18}$$

$$= \frac{4+4+5+5}{18} = 1$$

2.

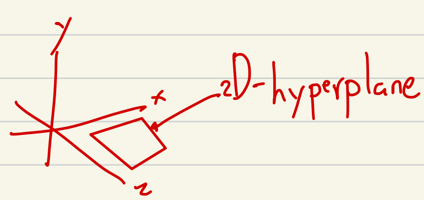
Figure 1.

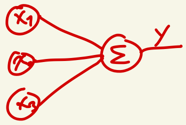
2. (2 marks) When the linear function has only **one** variable, show that its graph is a hyper-plane of dimension ***K*** (show *the linear equation of the graph*).

Linar Function with one variable $f(x_1) = a_1 x_1 + b$

the graph is hyperplane of dimension 1

h





X_1	X_2	X_3	Y
1	1	1	0
1	1	0	1
1	0	1	1
1	0	0	0
0	1	1	1
0	1	0	0
0	0	1	0
0	0	0	1

3. (1 mark) Check whether the following statement is **True** or **False**: In unsupervised learning, the Neural Network (NN) tries to "understand" and generate the output structure of the provided input data set "on its own" in the absence of its output pattern. ~~F~~ T? True

4. (1 mark) Check whether the following statement is **True** or **False**: Gradient descent is an iterative method to find the maximum of a function. F False

5. (1 mark) Check whether the following statement is **True** or **False**: Overfitting occurs if the model or algorithm shows high bias but low variance. F False

6.

or algorithm shows high bias but low variance. |

6. (2 marks) A three-input, one-output **parity detector** outputs a one if the number of "1" inputs is **even**; otherwise, it outputs a **0**. Can a Perceptron represent this function? If so, construct a Perceptron that does it; if not, argue why not.

A perceptron cannot represent parity detector since the output is linarly inseperable

7.

Perceptron that does it; if not, argue why not.

7. (5 marks) The Neural Network (NN) with its complete input, weight, and bias values are shown in **Figure 2**. Calculate the outputs of H1(2), H2(2), H3(2), and H1(3)out based on the following values: w11(1)= w12(1) = w13(1) = 0.4, w21(1) = w22(1) = w23(1) = 0.5, w31(1) = w32(1) = w33(1) = 0.3, w11(2) = w12(2) = w13(2) = 0.2, b1(1) = b2(1) = b3(1) = 0.7, b1(2) = 0.3, x1 = 2.5, x2 = 1.5, and x3 = 3.5. Use the **sigmoid** function as its output activation function.

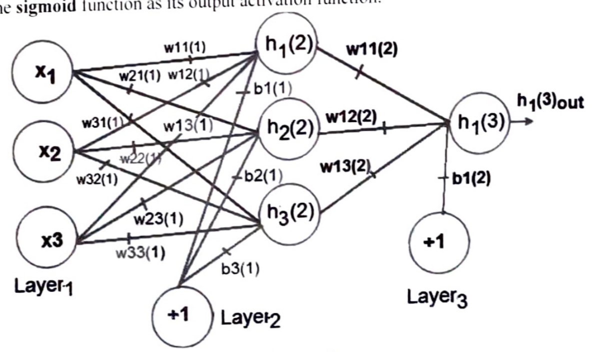


Figure 2.

a). $H_1^{(2)} = 0.4(2.5)+0.4(1.5)+0.4(3.5)+0.7 = 3.7$

$\frac{1}{1+e^{3.7}} = 0.9758$

b) $H_2^{(2)} = 0.5(2.5)+0.5(1.5)+0.5(3.5)+0.7 = 4.45$

$\frac{1}{1+e^{-4.45}} = 0.9885$

c) $H_3^{(2)} = 0.3(2.5)+0.3(1.5)+0.3(3.5)+0.7 = 2.95$

$\frac{1}{1+e^{-2.95}} = 0.9502$

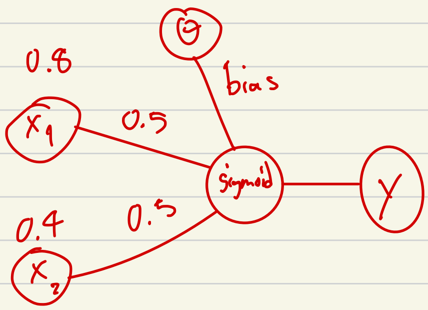
d) $H_1^{(3)} = 0.2(0.9758)+0.2(0.9885)+0.2(0.9502)+0.3 - 0.8829$

$\frac{1}{1+e^{-0.8829}} = 0.7074$

8. (4 marks) Consider a perceptron thawitho real-valued inputs and an output unit with a sigmoid activation function. All the initial weights and the bias(threshold) equal 0.5. Assume that the output should be 0.8 for the inputs x1 = 0.8 and x2 = -0.4. Show how the delta rule would alter this neural net upon processing (learning rate = 0.5). Show at least 1 iteration.

9. (3 marks) S...

X_1	X_2	Y_d (actual y)	W_1	W_2	Y (predict)	e	W1 new	W2 new
0.8	-0.4	0.8	0.5	0.5				



Y : $0.5(0.8) + 0.5(-0.4) - 0.5 = -0.3$

$y = \frac{1}{1+e^{-0.3}} = 0.43$

$e = Y_d - y = 0.37$

$W_1 \text{ new} = W_1 + \alpha(e)X_1$

$= 0.5 + 0.5(0.37)(0.8)$

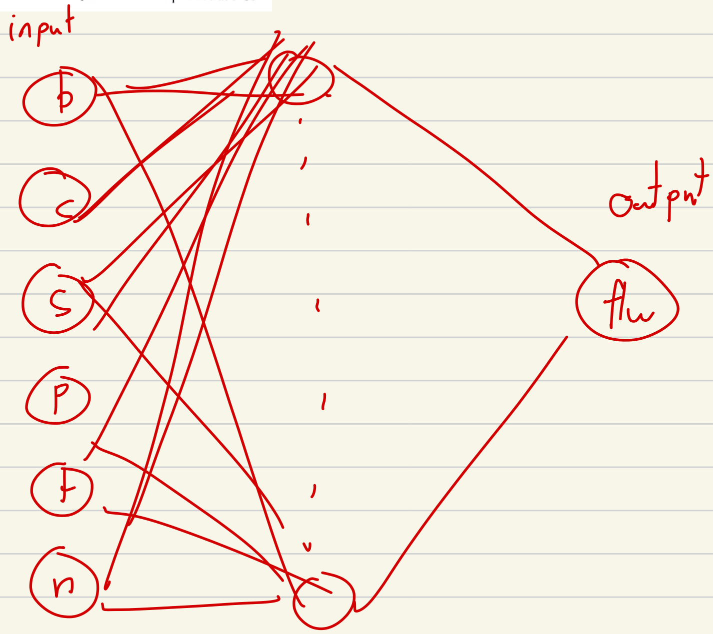
$= 0.648$

$W_2 \text{ new} = W_1 + \alpha(e)X_2$

$= 0.5 + 0.5(0.37)(-0.4)$

$= 0.43$

9. (3 marks) Suppose you use a backpropagation neural network (BPNN) for recognizing flu based on a patient dataset consisting of six input attributes (*body temperature*, *cough*, *body shivering*, *body pain*, *tastelessness*, and *nausea*) and ***one*** output value. All the values in the dataset are real numbers. Briefly describe how you will implement the prediction system (your description should include the topology of the BPNN, sample seen data, network training, and implementation procedure of the BPNN, etc.).



				Y
37.9	0.6	....	0.5	1
⋮				
36.9	0.8	. . . .	0.3	0

Weight [0.1, 0.2.... 0.3]

BackPNN

find new weight work back

10

the BPNN, etc.). ... and implementation procedure of

10. (3 marks) Calculate the following based on the training examples for a regression task shown in **Figure 3** (where X is input data, Y_d is desired output, Y is the predicted/calculated output, and E is the error value that must be calculated):

a) (1 mark) **0/1** Error.

b) (1 mark) The sum of squares error (SOSE).

c) (1 mark) Root means square error (RMSE).

Example	X	Yd	Y	E	0/1	e	e^2
e_1	0.7	1.7	1.7		0		
e_2	1.1	2.4	2.0		1	0.4	0.16
e_3	1.3	2.5	2.5		0		
e_4	1.9	1.5	1.5		0		
e_5	2.6	2.1	1.2		1	0.9	0.81
e_6	3.1	2.3	2.3		0		
e_7	3.9	3.2	2.8		1	0.4	0.16
e_8	2.9	1.8	1.8		0		
e_9	5.0	3.4	3.2		1	0.2	0.04

Figure 3.

11. (4 marks) Suppose a company wants to group the visitors to a website using just their age (one-

a. $0/1 \text{ Error} = \sum_{i \in 0/1} i = 1+1+1+1 = 4$

b. $SOSE = \sum e^2 = 0.16 + 0.81 + 0.16 + 0.04 = 0.117$

c. $RMSE = \sqrt{\frac{SOSE}{n}} = \sqrt{\frac{0.117}{9}} \approx 0.114$

11

11. (4 marks) Suppose a company wants to group the visitors to a website using just their age (one-dimensional space) as follows: 21, 22, 28, 35, 60, 61, 65, 40, 5, 15, 16, 19, 19, 20, 20, 41. Assume that the random initial centroids are **M1 = 16** and **M2 = 22**. Based on the given parameters, show the iterations (at least 2) used by the **k-means clustering** algorithm for **k = 2**. Same k=2

Iteration #1, $M_1 = 16$, $M_2 = 22$

$C_1 = \{5, 15, 16, 19, 19\}$

$C_2 = \{21, 22, 28, 35, 60, 61, 65, 40, 41\}$

New $M_1 = 17.44$ & $M_2 = 47.143$

Iteration #2

$C_1 = \{5, 15, 16, 19, 19, 21, 22, 28\}$

$C_2 = \{35, 60, 61, 65, 40, 41\}$

12

12. (3 marks) Based on the performance indicators of a Boolean predictor in **Figure 4**, calculate the following: a) [1 mark] *Precision*. b) [1 mark] *Recall*. c) [1 mark] *False-Positive Rate*.

	ap	*an*
pp	90	200
pn	50	500

Figure 4.

fp

$$\text{precision} = \frac{tp}{tp+fp} = \frac{90}{90+200}$$

fn

$$\text{recall} = \frac{tp}{tp+fn} \quad \frac{90}{90+50}$$

$$\text{fn-rate} = \frac{fn}{fn+tp} \quad \frac{200}{200+500}$$

13

Figure 4.

13. (5 marks) Suppose a spinning wheel game with nine numbers [1,2,3,4,5,6,7,8,9], each with equal probability. Let S be the outcome of a spin. Based on these:

a) (2 marks) Find the information (entropy) **H(*S*)** of a spin.

b) (2 marks) Suppose an **odd sensor *O*** is connected and detects only the odd value from a spin. Then, find **H(*S*|*O*)**.

c) (1 mark) Find the **information gain (IG)** of the spin with the odd sensor ***O***.

a) $$H(S) = -\sum_{x=1}^{9} P(X=x)\log_2 P(X=x)$$

$$= -\sum_{x=1}^{9} \tfrac{1}{9}\log_2(\tfrac{1}{9})$$

$$= -9 \times \tfrac{1}{9} \times \frac{\log_{10}(1/9)}{\log_{10}(2)}$$

$$= -(-3.167) = 3.167$$

b) $$H(S|O) = -\frac{5}{9}\log_2 \tfrac{1}{5} - \tfrac{4}{9}\log_2 \tfrac{1}{4}$$

$$= 2.197$$

c) $$IG = H(S) - H(S|O)$$

14.

14. (4 marks) You believe **1%** of sports members in your group use steroids. You have arranged a blood test that detects steroids: There is a **99%** chance you will be caught positive to steroids when you use steroids. There is a **0.5%** chance you will be positive if you do not use steroids. What is the chance that a positive blood analysis identifies a steroid user? (Hints: Let ***A*** is "*use steroids*," and ***B*** is "*positive to steroids*"). Show all of your calculation steps.

$P(A|B)$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\sim A)P(\sim A)}$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$P(S) = 0.01$

$P(+|S) = 0.99$

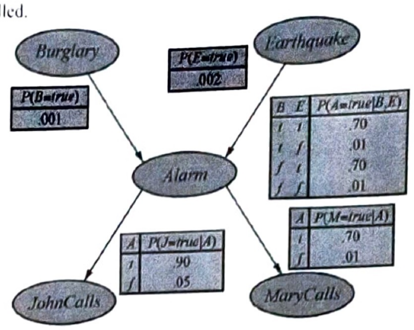
$P(+|\sim S) = 0.005$

$$P(S|+) = \frac{P(+|S)P(S)}{P(+|S)P(S) + P(+|\sim S)P(\sim S)}$$

$$= \frac{0.99(0.01)}{0.99(0.01) + (0.005)(0.99)}$$

15

15. (3 marks) Calculate the probability of the following event based on the Bayes net shown in **Figure 5**: "The alarm (***A***) has sounded, but no burglary (***B***) has occurred, but a minor earthquake (***E***) has occurred, and John (***J***) called and Mary (***M***) not called.



$P(A \wedge \sim B \wedge E \wedge J \wedge \sim M) = P(\sim B) \times P(E) \times P(A|\sim B \wedge E) \times P(J|A) \times P(\sim M|A)$

$= 0.99 \times 0.02 \times 0.7 \times 0.9 \times 0.3$

16

Figure 5.

16. (5 marks) The joint probability distribution of three variables, flu (f), allergy(a), and sinus (s), is shown in **Figure 6**. By applying the direct computation of the joint probability distribution, check whether $\mathbf{P(f^1|s^1) < P(f^1|a^1)}$.

1	f^1	a^1	s^1	0.0270
2	f^1	a^1	s^0	0.0030
3	f^1	a^0	s^1	0.1620
4	f^1	a^0	s^0	0.1080
5	f^0	a^1	s^1	0.0140
6	f^0	a^1	s^0	0.0560
7	f^0	a^0	s^1	0.0063
8	f^0	a^0	s^0	0.6237

Figure 6.

17. (6 marks) Consider the Bayesian Network with three Boolean variables shown in **Figure 7.**

$$P(f^1|s^1) = \frac{P(f^1 \cap s^1)}{P(s^1)} = \frac{①+③}{①+③+⑤+⑦} = \frac{0.189}{0.189+0.0203} = \frac{0.189}{0.2093} = 0.903$$

$$P(f^1|a^1) = \frac{P(f^1 \cap a^1)}{P(a^1)} = \frac{①+②}{①+②+⑤+⑥} = \frac{0.03}{0.03+0.07} = \frac{0.03}{0.1} = 0.3$$

QED

Figure 6.

17 17. (6 marks) Consider the Bayesian Network with three Boolean variables shown in **Figure 7.** Compute a) [2 marks]. P(~A | W, H). b) [2 marks] P(~A, W, H). c) [2 marks] P(~A | H).

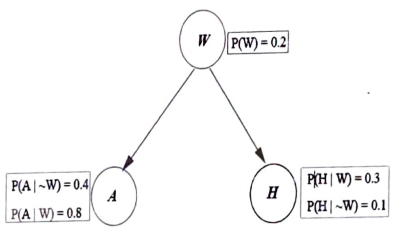


Figure 7.

a. $P(\sim A | W,H) = \frac{P(\sim A \wedge W \wedge H)}{P(W \wedge H)} = \frac{0.2 \times 0.2 \times 0.3}{0.2 \times 0.3} = 0.2$

b $P(\sim A \wedge W \wedge H) = P(W) \times P(\sim A|W) \times P(H|W)$

$= 0.2 \times 0.2 \times 0.3$

c. $P(\sim A \mid H) = \dfrac{P(\sim A \wedge H)}{P(H)} = \dfrac{P(H \wedge \sim A \wedge W) + P(H \wedge \sim A \wedge \sim W)}{P(H \wedge \sim A \wedge W) + P(H \wedge \sim A \wedge \sim W) + P(H \wedge A \wedge W) + P(H \wedge A \wedge \sim W)}$

18.

Figure 7.

18. (4 marks) **Figure 8** shows the two factors, ***go_out***, and ***get_coffee***, from a probabilistic inference function. Show **summing out** the variable ***Rain*** from the product of ***get_coffee*** and ***go_out***.

get_coffee

Full	*Wet*	*Prob*
t	*t*	0.6
t	*f*	0.4
f	*t*	0.3
f	*f*	0.7

go_out

Rain	*Wet*	*Prob*
t	*t*	0.8
t	*f*	0.2
f	*t*	0.1
f	*f*	0.9

Figure 8.

1. times & Natural Join

Full	Wet	Rain	Prob
t	t	t	0.48
t	t	f	0.06
t	f	t	0.08
t	f	f	0.36
f	t	t	0.24
f	t	f	0.03
f	f	t	0.14
f	f	f	0.63

2. Summing

Full	Wet	Prob
t	t	0.54
t	f	0.44
f	t	0.27
f	f	0.77

e.

4 $P(D) = P(D|C)P(C) + P(D|\sim C)P(\sim C)$

$$P(C) = P(C \wedge A \wedge B) + P(C \wedge A \wedge \sim B) + P(C \wedge \sim A \wedge B) + P(C \wedge \sim A \wedge \sim B)$$

$$= P(C|A,B)P(A \wedge B) + P(C|A,\sim B)P(A \wedge \sim B) + P(C|\sim A \wedge B)P(\sim A \wedge B)$$

$$+ P(C|\sim A,\sim B)P(\sim A \wedge \sim B)$$

Assumption University
Vincent Mary School of Science and Technology
CS3425/CSX4209/ITX4209 ISD

Quiz II Semester 2/2022 (60 marks)

Instructions: Students can use an electronic calculator and a ***black*** or ***blue*** pen for answering. This is a **closed-book** examination.

1. (4 marks) Consider the supervised learning dataset in **Figure 1**. Calculate the **log loss** of splitting the input attribute "***WhereRead***" in '***UserAction***' (Show all your calculation steps).

Example	*Author*	*Thread*	*Length*	*WhereRead*	*UserAction*
e_1	known	new	long	home	skips
e_2	unknown	new	short	work	reads
e_3	unknown	follow Up	long	work	skips
e_4	known	follow Up	long	home	skips
e_5	known	new	short	home	reads
e_6	known	follow Up	long	work	skips
e_7	unknown	follow Up	short	work	skips
e_8	unknown	new	short	work	reads
e_9	known	follow Up	long	home	skips
e_{10}	known	new	long	work	skips
e_{11}	unknown	follow Up	short	home	skips
e_{12}	known	new	long	work	skips
e_{13}	known	follow Up	short	home	reads
e_{14}	known	new	short	work	reads
e_{15}	known	new	short	home	reads
e_{16}	known	follow Up	short	work	reads
e_{17}	known	new	short	home	reads
e_{18}	unknown	new	short	work	reads
e_{19}	unknown	new	long	work	? Unseen
e_{20}	unknown	follow Up	long	home	?

Figure 1.

2. (2 marks) When the linear function has only **one** variable, show that its graph is a hyper-plane of dimension ***K*** (show *the linear equation of the graph*).
3. (1 mark) Check whether the following statement is **True** or **False**: In unsupervised learning, the Neural Network (NN) tries to "understand" and generate the output structure of the provided input data set "on its own" in the absence of its output pattern. K T ?
4. (1 mark) Check whether the following statement is **True** or **False**: Gradient descent is an iterative method to find the maximum of a function. F

Ch.7 p.70

5. (1 mark) Check whether the following statement is **True** or **False**: Overfitting occurs if the model or algorithm shows high bias but low variance. F
6. (2 marks) A three-input, one-output **parity detector** outputs a one if the number of "1" inputs is **even**; otherwise, it outputs a **0**. Can a Perceptron represent this function? If so, construct a Perceptron that does it; if not, argue why not.
7. (5 marks) The Neural Network (NN) with its complete input, weight, and bias values are shown in **Figure 2**. Calculate the outputs of H**1(2)**, H**2(2)**, H**3(2)**, and H**1(3)$_{out}$** based on the following values: w11(1)= w12(1) = w13(1) = 0.4, w21(1) = w22(1) = w23(1) = 0.5, w31(1) =w32(1) = w33(1) = 0.3 w11(2) =w12(2) = w13(2) = 0.2,b1(1) = b2(1) = b3(1) = 0.7, b1(2) =0.3, x1 = 2.5, x2 = 1.5, and x3 = 3.5. Use the **sigmoid** function as its output activation function.

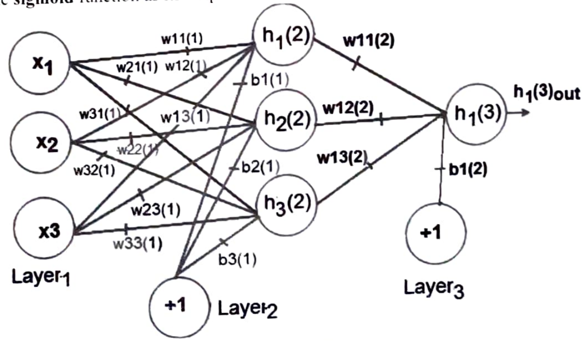


Figure 2.

$\frac{1}{1+e^{-z}}$

8. (4 marks) Consider a perceptron thawitho real-valued inputs and an output unit with a sigmoid activation function. All the initial weights and the bias(threshold) equal 0.5. Assume that the output should be 0.8 for the inputs $x1 = 0.8$ and $x2 = -0.4$. Show how the delta rule would alter this neural net upon processing (learning rate = 0.5). Show at least 1 iteration.
9. (3 marks) Suppose you use a backpropagation neural network (BPNN) for recognizing flu based on a patient dataset consisting of six input attributes (*body temperature, cough, body shivering, body pain, tastelessness*, and *nausea*) and ***one*** output value. All the values in the dataset are real numbers. Briefly describe how you will implement the prediction system (your description should include the topology of the BPNN, sample seen data, network training, and implementation procedure of the BPNN, etc.).
10. (3 marks) Calculate the following based on the training examples for a regression task shown in **Figure 3** (where ***X*** is input data, Y_d is desired output, ***Y*** is the predicted/calculated output, and ***E*** is the error value that must be calculated):
    a) (1 mark) **0/1** Error.
    b) (1 mark) The sum of squares error (SOSE).
    c) (1 mark) Root means square error (RMSE).

Example	X	Yd	Y	E	
e_1	0.7	1.7	1.7		0
e_2	1.1	2.4	2.0		1
e_3	1.3	2.5	2.5		0
e_4	1.9	1.5	1.5		0
e_5	2.6	2.1	1.2		1
e_6	3.1	2.3	2.3		0
e_7	3.9	3.2	2.8		1
e_8	2.9	1.8	1.8		0
e_9	5.0	3.4	3.2		1

Figure 3.

11. (4 marks) Suppose a company wants to group the visitors to a website using just their age (one-dimensional space) as follows: 21, 22, 28, 35, 60, 61, 65, 40, 5, 15, 16,19, 19, 20, 20, 41. Assume that the random initial centroids are **M1 = 16** and **M2 = 22**. Based on the given parameters, show the iterations (at least 2) used by the **k-means clustering** algorithm for **k** = 2. Same k=2
12. (3 marks) Based on the performance indicators of a Boolean predictor in **Figure 4,** calculate the following: a) [1 mark] *Precision*. b) [1 mark] *Recall*. c) [1 mark] *False-Positive Rate*.

	ap	*an*
pp	90	200
pn	50	500

Figure 4.

13. (5 marks) Suppose a spinning wheel game with nine numbers [1,2,3,4,5,6,7,8,9], each with equal probability. Let ***S*** be the outcome of a spin. Based on these:
    a) (2 marks) Find the information (entropy) **H(*S*)** of a spin.
    b) (2 marks) Suppose an **odd sensor *O*** is connected and detects only the odd value from a spin. Then, find **H(*S*|*O*)**.
    c) (1 mark) Find the **information gain (IG)** of the spin with the odd sensor ***O***.
14. (4 marks) You believe **1%** of sports members in your group use steroids. You have arranged a blood test that detects steroids: There is a **99%** chance you will be caught positive to steroids when you use steroids. There is a **0.5%** chance you will be positive if you do not use steroids. What is the chance that a positive blood analysis identifies a steroid user? (Hints: Let ***A*** is "*use steroids*," and ***B*** is "*positive to steroids*"). Show all of your calculation steps.

P(A|B)

15. (3 marks) Calculate the probability of the following event based on the Bayes net shown in **Figure 5**: "The alarm (*A*) has sounded, but no burglary (*B*) has occurred, but a minor earthquake (*E*) has occurred, and John (*J*) called and Mary (*M*) not called.

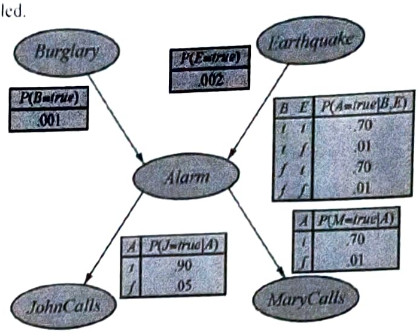


Figure 5.

16. (5 marks) The joint probability distribution of three variables, flu (*f*), allergy(*a*), and sinus (*s*), is shown in **Figure 6**. By applying the direct computation of the joint probability distribution, check whether $\mathbf{P(f^1|s^1) < P(f^1|a^1)}$.

f^1	a^1	s^1	0.0270
f^1	a^1	s^0	0.0030
f^1	a^0	s^1	0.1620
f^1	a^0	s^0	0.1080
f^0	a^1	s^1	0.0140
f^0	a^1	s^0	0.0560
f^0	a^0	s^1	0.0063
f^0	a^0	s^0	0.6237

Figure 6.

17. (6 marks) Consider the Bayesian Network with three Boolean variables shown in **Figure 7**. Compute a) [2 marks]. P(~A | W, H). b) [2 marks] P(~A, W, H). c) [2 marks] P(~A | H).

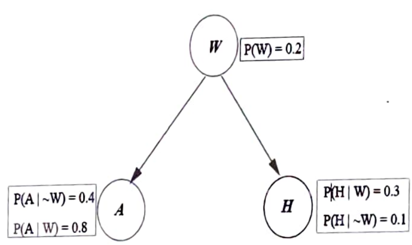


Figure 7.

18. (4 marks) **Figure 8** shows the two factors, ***go_out***, and ***get_coffee***, from a probabilistic inference function. Show **summing out** the variable ***Rain*** from the product of ***get_coffee*** and ***go_out***.

get_coffee

Full	*Wet*	*Prob*
t	*t*	0.6
t	*f*	0.4
f	*t*	0.3
f	*f*	0.7

go_out

Rain	*Wet*	*Prob*
t	*t*	0.8
t	*f*	0.2
f	*t*	0.1
f	*f*	0.9

Figure 8.