ASSUMPTION UNIVERSITY
VINCENT MARY SCHOOL OF ENGINEERING, SCIENCES, AND TECHNOLOGY
CSX4209/CSX4281/ITX4209 QUIZ 2 KEY (3 hrs. 58 points) 1/2024

1. (5 points) You believe **1%** of gym 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 for 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*** be "*use steroids*," and ***B*** be "*positive to steroids*"). Show all your calculation steps.

   P(A) = 0.01 (1%)
   P(~A) = 1 – P(A) = **1 – 0.01 = 0.99**
   P(B|A) = 0.99 (99%)
   P(~B|A) = 1 – 0.99 = 0.01
   P(B|~A) = 0.005 (0.5%)
   Find P(A|B)?
   P(A|B) = P(B|A) × P(A)/P(B)
   where P(B) = P(B|A) × P(A) + P(B|~A) × P(~A)
   P(A|B) = (0.99 × 0.01) / (0.99 × 0.01 + 0.005 × 0.99) = 0.667 = 66.7%

2. (4 points) Calculate the following based on the training examples for a regression task shown in **Figure 1 below** (where ***X*** is input data, Y_d is desired output, and ***Y*** is the predicted/calculated output): **a)** Sum of **0/1** Error = **4**. **b)** Sum of squares error (SOSE) = **1.17**. **c)** Root means square error (RMSE) = **0.361**.

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

3. (3 points) Based on the performance indicators of a Boolean predictor in **Figure 2 below,** calculate the following: a) *Precision*, b) *Recall*, and c) *False-Positive Rate.*

	ap	*an*
pp	90	200
pn	50	500

a) [1 mark] **Precision** = **tp/(tp+fp)** = 90/(90+200) = **0.3103**
b) [1 mark] **Recall (True-positive rate)** = **tp/(tp+fn)** = 90/(90+50) = **0.6429**
c) [1 mark] **False-positive rate** = **fp/(fp+tn)** = 200/(200+500) = **0.2857**

4. (5 points) Suppose a company wants to group their visitors by their age. The ages of the first set of visitors are {22, 46, 18, 35, 60, 24, 65, 40, 55, 39, 50}. Assume that the initial centroids are **M1** = 18 and **M2** = 60. Based on the given parameters, show the results of the first iteration used by the ***k*-means clustering** algorithm for ***k*** = **2** by filling the blanks of the following table (where D1 and D2 are the distances of data from their centroids).

NAME .... **SURNAME** .... **ID.NO.** .... **SEC.**........

Data	D1	D2	Cluster
22	4	38	C1
46	28	14	C2
18	0	42	C1
35	17	25	C1
60	42	0	C2
24	6	36	C1
65	47	5	C2
40	22	20	C2
55	37	5	C2
39	21	21	C1
50	32	10	C2

5. (5 points) Consider the supervised learning dataset in **Figure 3 below.** Calculate the **log loss** of the input attribute "***WhereRead***" in the output attribute '***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	?

<u>**ANS:**</u> Splitting on ***WhereRead*** divides the examples in ***UserAction*** into ***home*** → (4 ***skips***) and (4 ***reads***). Similarly, ***work*** → (5 ***skips***) and (5 ***reads***). Therefore, <u>**four**</u> combinations of entropy values need to be calculated:

(i). Number of ***skips*** under ***home*** in ***UserAction*** × **log₂** (probability of ***skips*** under ***home*** in ***UserAction***) = **4 ×** $\mathbf{\log_2}$ **(4 *skips*)/(4 *skips* + 4 *reads*) = 4 ×** $\mathbf{\log_2(4/8)}$ **= 4 ×** $\mathbf{\log_2 0.5}$

(ii). Number of ***reads*** under ***home*** in ***UserAction*** × $\mathbf{\log_2}$ (probability of ***reads*** under ***home*** in ***UserAction***) = **4 ×** $\mathbf{\log_2}$ **(4 *reads*)/(4 *reads* + 4 *skips*) = 4 ×** $\mathbf{\log_2(4/8)}$ **= 4 ×** $\mathbf{\log_2 0.5}$

(iii). Number of ***skips*** under ***work*** in ***UserAction*** × $\mathbf{\log_2}$ (probability of ***skips*** under ***work*** in ***UserAction***) = **5 ×** $\mathbf{\log_2}$ **(5 *skips*)/(5 *reads* + 5 *skips*) = 5 ×** $\mathbf{\log_2(5/10)}$ **= 5 ×** $\mathbf{\log_2 0.5}$

(iv). Number of ***reads*** under ***work*** in ***UserAction*** × $\mathbf{\log_2}$ (probability of ***reads*** under ***work*** in ***UserAction***) = **5 ×** $\mathbf{\log_2}$ **(5 *reads*)/(5 *reads* + 5 *skips*) = 5 ×** $\mathbf{\log_2(5/10)}$ **= 5 ×** $\mathbf{\log_2 0.5}$

NAME ……………………………… SURNAME ……………………………… ID.NO. ………… SEC………

The log loss = -[(4 × $\log_2$0.5) + (4 × $\log_2$0.5) + (5 × $\log_2$0.5) +(5 × $\log_2$0.5)]/18

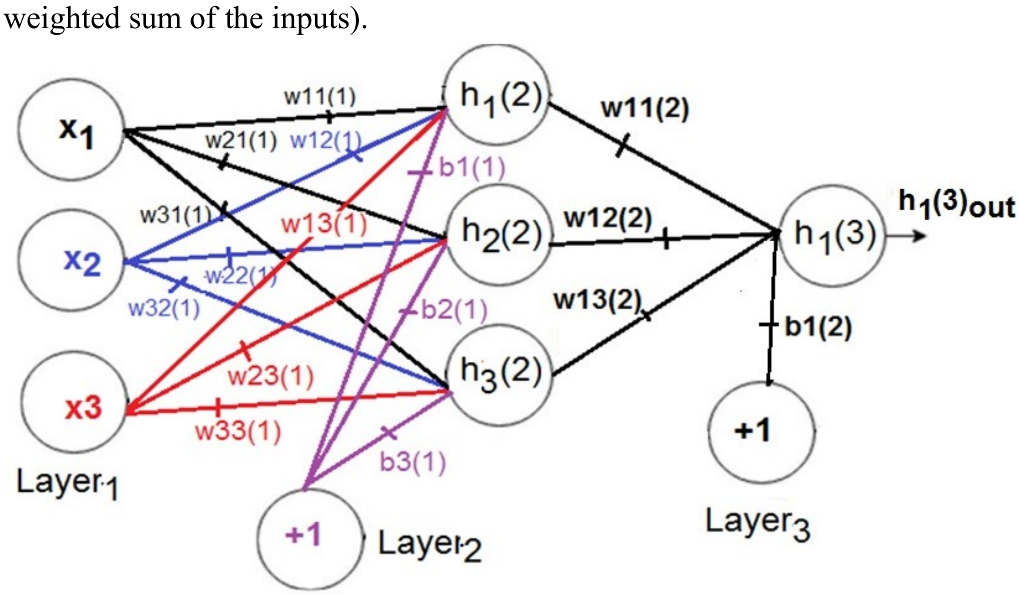
Where $\log_2$ 0.5 = -1

<u>So the log loss</u> = (4 + 4 + 5 + 5)/18 = <u>1</u>

6. (5 points) A section of **perceptron training** for the **2-input OR** function is shown in the following **Table** (where **w1**, **w2**, and **y** are the weights, and the calculated output of the perceptron and **x1**, **x2**, and **yd** are the inputs and output of the **OR** function). Assume the **w1= -0.4, w2 = 0.3**, **bias** (**b**) = **0.2,** and **learning rate** (∝) = **0.1**. Show the perceptron's first iteration by completing the table's blanks (the <u>**step** activation</u> estimates **y,** and the <u>**bias is subtracted**</u> from the weighted sum of inputs).

x1	x2	yd	w1	w2	y	e	w1new	w2new
0	0	0	-0.4	0.3	0	0	-0.4	0.3
1	0	1	-0.4	0.3	0	1	-0.3	0.3
0	1	1	-0.3	0.3	1	0	-0.3	0.3
1	1	1	-0.4	0.3	0	1	-0.3	0.4

7. (8 points) The **multi-layer perceptron (MLP)** with its input, weight, and bias values are shown in **Figure 5**. Calculate its feed-forward 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** (the <u>**sigmoid** activation</u> estimates output of **neurons** and the <u>**bias is added**</u> to the weighted sum of the inputs).



a) <u>***h*1(2):**</u>
X1(2) = x1w11(1) + x2w31(1) + x3w13(1) + b1(1)
<u>Output, h1(2)</u> = 1/(1 + $e^{-X1(2)}$) = <u>0.9723</u>

b) <u>***h*2(2):**</u>
X2(2) = x1w21(1) + x2w22(1) + x3w23(1) + b2(1)
<u>Output, h2(2)</u> = 1/(1 + $e^{-X2(2)}$) = <u>0.9885</u>

c) <u>***h*3(2):**</u>
X3(2) = x1w31(1) + x2w32(1) + x3w33(1) + b3(1)
<u>Output, h3(2)</u> = 1/(1 + $e^{-X3(2)}$) = <u>0.9504</u>

d) <u>***h*1(3)out:**</u>
X3 = h1(2)w11(2) + h2(2)w12(2) + h3(2)w13(2) + b1(2)
<u>Output, *h*1(3)out</u> = 1/(1 + e^{-X3}) = <u>0.6976</u>

NAME ……………………………… **SURNAME** ……………………………… **ID.NO.** ………… **SEC.**………

8. (5 points) 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 points) Find the information (entropy) **H(*S*)** of a spin. **(b).** (2 points) Suppose an **odd sensor *O*** is connected and detects only the odd value from a spin. Find **H(*S*|*O*)** **(c).** (1 point). Find the **spin's information gain (IG)** with the odd sensor ***O***.

a) (2 points) The 9 values on the spinning wheel are {1,2,3,4,5,6,7,8,9}.

Let S bet the outcome of a spin. Then the Entropy of S, H(S) is given as:

H(S) $= - \sum_{i=1}^{9} 1/9 \times \log_2 1/9$

$= - \sum_{i=1}^{9} 0.111 \times \log_2 0.111$

$= \underline{3.17}$

b) (2 points) The odd sensor O, detects one value out of {1,3,5,7,9} and never detects any even value {2,4,6,8}, and it skips these even values.

The entropy of a spin in the presence of the odd sensor O is H(S|O):

H(S|O) = – (5 possible odd outcome/total values) × $\log_2$ (one odd outcome)/5 odd values) – (4 blocked even values/total values) × $\log_2$ (one even outcome/ 4 blocked even values)

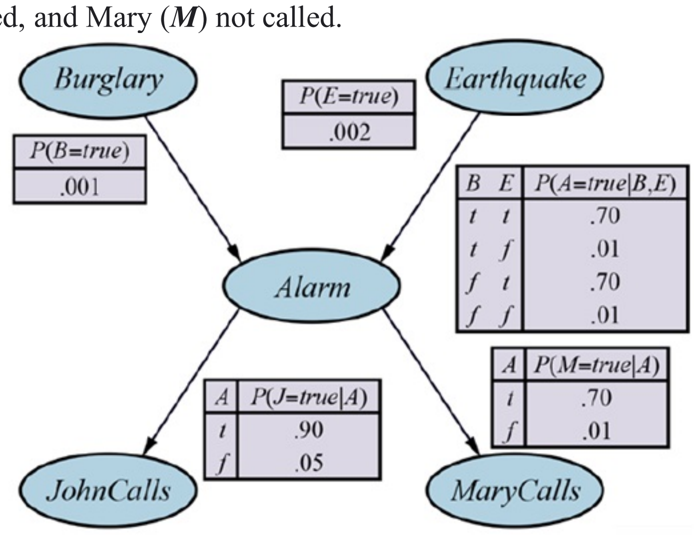
= - 5/9 × $\log_2$ (1/5) – 4/9 × $\log_2$ (¼)

= - 0.5556 × $\log_2$ 0.2 – 0.444 × $\log_2$ 0.25

<u>= 2.179</u>

c) (1 point) Information gain (IG) = H(S) – H(S|O) = 3.17 – 2.179 = <u>0.991</u>

9. (3 points) Calculate the joint probability of the following event based on the Bayes net shown in **Figure below**: "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 ∧ ~B ∧ E ∧ J ∧ ~M)

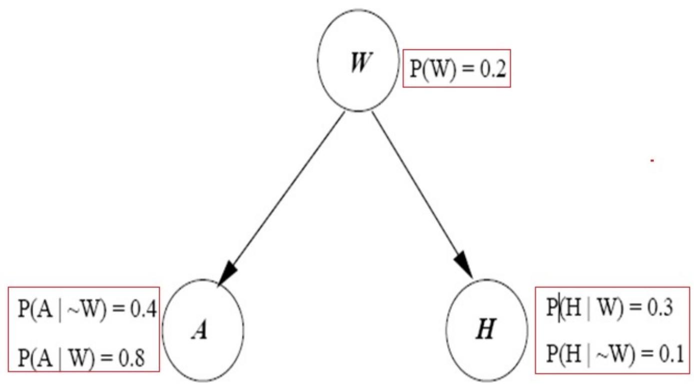
= P(A|~B∧E) × P(~B) × P(E) × P(J|A) × P(~M|A)

= 0.7 × (1-0.001) × 0.002 × 0.90 × (1-0.70)

= 0.7 × 0.999 × 0.002 × 0.90 × 0.30 = <u>0.000378</u>

NAME ……………………………… SURNAME ……………………………… ID.NO. ………… SEC………

10. (4 points) Consider the Bayesian Network with three Boolean variables shown in **Figure below.** Compute **(a)** (2 points) P(~A | W, H). **(b)** (2 points) P(~A, W, H).



a) (2 points) P(~A|W, H) = P(~A|W∧ H)
= P(~A|W) = 1 – P(A|W) = 1 – 0.8 = <u>0.2</u>

b) (2 points) P(~A, W, H) = P(~A ∧ W ∧ H)
= P(~A|W) × P(W) × P(H|W)
= 0.2 × 0.2 × 0.3 = <u>0.012</u>

11. (6 points) The joint probability distribution of three variables, flu (***f***), allergy(***a***), and sinus (***s***), is shown in **Figure below**. By applying the direct computation of the joint probability distribution, check whether **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

<u>P(f|s) = P(f ∧ s)/P(s)</u>
Where P(f ∧ s) = P(f ∧ a ∧ s) + P(f ∧ ~a ∧ s) = 0.0270 + 0.1620 = 0.189
Similarly, P(s) = P(f ∧ a ∧ s) + P(~f ∧ a ∧ s) + P(f ∧ ~a ∧ s) + P(~f ∧ ~a ∧ s)
= 0.027 + 0.014 + 0.1620 + 0.0063 = 0.2093
P(f|s) = P(f ∧ s)/P(s) = 0.189/0.2093 = <u>0.903</u>
<u>P(f|a) = P(f ∧ a)/P(a)</u>
Where P(f ∧ a) = (f ∧ a ∧ s) + P(f ∧ a ∧ ~s) = 0.0270 + 0.0030 = 0.03
Similarly, P(a) = P(f ∧ a ∧ s) + P(~f ∧ a ∧ s) + P(f ∧ a ∧ ~s) + P(~f ∧ a ∧ ~s)
= 0.0270 + 0.0140 + 0.0030 + 0.0560 = 0.10
P(f|a) = P(f ∧ a)/P(a) = 0.03/0.10 = <u>0.3</u> Therefore, *P*(f^1|s^1) < *P*(f^1|a^1) is proved.

12. (5 points) The **Figure below** shows the two factors, ***go_out,*** and ***get_coffee***, from a probabilistic inference function. Show how the **summing-out** process eliminates 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

NAME ……………………………… SURNAME ……………………………… ID.NO. ………… SEC………

Step 1: Find the product of factors *get_coffee* and *go_out*
Assume that, *Full* = *F*, *Wet* = *W*, *Rain* = *R*

F	*W*	*R*	Probability
t	*t*	*t*	0.48 (0.6 × 0.8)
t	*t*	*f*	0.06 (0.6 × 0.1)
t	*f*	*t*	0.08 (0.4 × 0.2)
t	*f*	*f*	0.36 (0.4 × 0.9)
f	*t*	*t*	0.24 (0.3 × 0.8)
f	*t*	*f*	0.03 (0.3 × 0.1)
f	*f*	*t*	0.14 (0.7 × 0.2)
f	*f*	*f*	0.63 (0.7 × 0.9)

Step 2: From the product of the factors *get_coffee* and *go_out*, apply the summing technique to eliminate the variable *Rain(R)*.

F	*W*	Probability
t	*t*	0.54
t	*f*	0.44
f	*t*	0.27
f	*f*	0.77

$$\left.\begin{array}{l} F = t,\ W = t,\ R = t = 0.48 \\ F = t,\ W = t,\ R = f = 0.06 \end{array}\right\} \quad 0.48 + 0.06 = 0.54 \qquad \text{i.}$$

$$\left.\begin{array}{l} F = t,\ W = f,\ R = t = 0.08 \\ F = t,\ W = f,\ R = f = 0.36 \end{array}\right\} \quad 0.08 + 0.36 = 0.44 \qquad \text{ii.}$$

$$\left.\begin{array}{l} F = f,\ W = t,\ R = t = 0.24 \\ F = f,\ W = t,\ R = f = 0.03 \end{array}\right\} \quad 0.24 + 0.03 = 0.27 \qquad \text{iii.}$$

$$\left.\begin{array}{l} F = f,\ W = f,\ R = t = 0.14 \\ F = f,\ W = f,\ R = f = 0.63 \end{array}\right\} \quad 0.14 + 0.63 = 0.77 \qquad \text{iv.}$$