[17%] True/False Questions

ML and Decision Tree:

1- In decision tree learning, the max-gain rule chooses an attribute that can split a set of examples into new subsets that have the highest entropy. (False)

2- Bias is an assumption for how big the search space for the target function is. (True)

3- Entropy is a measure of the effectiveness of a branch in a decision tree. (True)

NN and Deep Learning:
4- Back-propagation is changing the weights in the network based on the output errors. (True)

5- The learning that occurs in a NN are the adjustments of the weights inside the NN（T）

Probability:

6- If the existing values of a random variable are not exhaustive, then adding a new (default) value will make the new set of values exhaustive. (T)

7- If A and B are mutually exclusive, then they are not independent. （T）

8- If A -> B, then B depends on A and A is independent of B. (False, the 2nd part)

Bayesian Networks:

9- Given a fully joint probability distribution table, we cannot do all probabilistic reasonings without any Bayesian Networks. (f)

10- The size of a Bayesian Network is usually smaller than its corresponding fully joint probability distribution table. （T）

Maximal Expected Utility (MEU) and Rationality

11- If an agent follows the principle of MEU, then it behaves rationally. (t)

12- Expected Utilities can be used by an agent to select actions intelligently. (T)

Bayesian Learning/Classifier

13- Naïve Bayesian Classifier can only be used when random variables are assumed to be mutually independent. (T)

14- Bayesian Learner/Classifier can deal with uncertainties and noise in the training data. (T)

Enumeration algorithm:

15- The enumeration algorithm uses the only the product rule. (F)

Approximation algorithms (approach to the real probability)

16- All approximation reasoning algorithms of BBnets should be provable to approach the true probability when the sample size goes to infinity. (T)

Hidden Markov Models

17- If the current state distribution is known, then the previous observation history of the HMM is critical for reasoning for the future. (f)

[14%] Decision Tree Learning

You are given the task of classifying the electrical consumption of a household based on 3 attributes **[High use of electrical appliances, Location of the house, Type of family]** as shown in the following table :

As you can observe - “high use of electrical appliances” can be yes or no, “location” of the house can be Coastal or Interior and the “type of family” can be joint or nuclear. The data provided is fictitious and for the purpose of testing your knowledge about decision trees only.

Index	High Use of Appliances	Location of the house	Type of Family	Electricity Consumption
1	yes	coastal	nuclear	***low***
2	yes	coastal	joint	***low***
3	no	coastal	nuclear	***low***
4	yes	interior	nuclear	***high***
5	no	coastal	joint	***high***
6	yes	interior	joint	***high***
7	no	interior	nuclear	***low***
8	no	interior	joint	***high***

Answer the following questions based on the information provided above :

The following values are provided for your reference -

0 * log (0) = 0
1/4 * log (1/4) = - 0.5
2/4 * log (2/4) = - 0.5
3/4 * log (3/4) = -0.3
4/4 * log (4/4) = 0

Please note we would only be grading based on log base 2 values (using which the above values were calculated). Please do not use any other log base.

[1%][A] Calculate the entropy of the decision [Electricity Consumption]

Entropy (Electricity Consumption) = [-0.5 * log(0.5) - 0.5 log(0.5)] = 1

For B,C,D partial marking - 1% for correct formula application/correct raw values, 1% for correct final answer

[2%][B] Calculate the entropy remaining (remainder) if "high use of appliances" is chosen as the splitting attribute.

Entropy (high use of appliances) =
½ * [- 1/2 * log (1/2) - 1/2 * log(1/2)] + ½ * [- 1/2 * log(1/2) - 1/2 * log(1/2)] = 1

[2%][C] Calculate the entropy remaining (remainder) if "location of the house" is chosen as the splitting attribute.

Entropy (location of the house) = ½ * [- 3/4 * log(3/4) - 1/4 * log(1/4)] + ½ * [- 1/4 * log (1/4) - 3/4 * log(3/4)] = 0.5 * (0.3+0.5) + 0.5 * (0.3+0.5) = 0.8

[2%][D] Calculate the entropy remaining (remainder) if "type of family" is chosen as the splitting attribute.

Entropy (type of family) = ½ * [- 3/4 * log(3/4) - 1/4 * log(1/4)] + ½ * [- 1/4 * log (1/4) - 3/4 * log(3/4)] = 0.5 * (0.3+0.5) + 0.5 * (0.3+0.5) = 0.8

[3%][E] Calculate the information gain for each attribute

InfoGain (high use of appliances) =
InfoGain (location of the house) =
InfoGain (type of family) =

InfoGain (high use of appliances) = **1 - 1 = 0 [1%]**
InfoGain (location of the house) = **1 - 0.8 = 0.2 [1%]**
InfoGain (type of family) = **1 - 0.8 = 0.2 [1%]**

[2%][F] Which attribute should NOT be chosen to split the dataset and why?
In case of a tie, please mention all such attributes.

High use of Appliances[1%]. Reason : It has lowest information gain.[1%]

[2%][G] How many distinct decision trees are possible with n Boolean attributes?

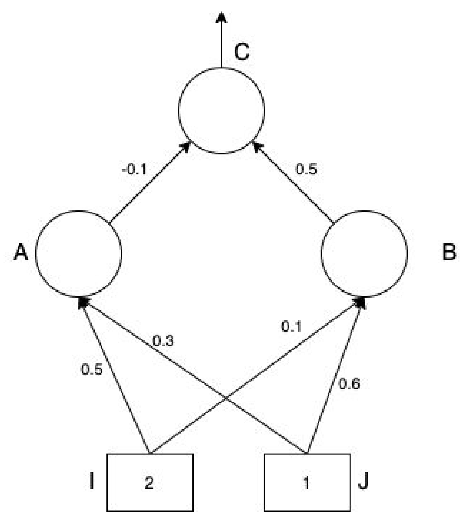
1. n
2. 2 ^ n
3. 2 ^ (2 ^ n)
4. n!

[Option 3]

[14%] Neural Networks

1.[8%] Consider the neural network shown below, where I, J are inputs and A, B, C are perceptrons. For each perceptron assume it's firing is based on the threshold rule discussed in class with Threshold = 1 i.e.

$$f(x) = \begin{cases} 1, & if\ x \geq 1 \\ 0, & otherwise \end{cases}$$



1.1[3%] Write the output for each perceptron A,B and C.

A = 1
B = 0
C = 0
[1% for each, no partial points]

1.2[5%] Given the expected output to be 1, find the error value, and the updated weights for AC and BC. Use the value of alpha as 1.

a. [1%] Expected Error =
b. [2%] AC_new =
c. [2%] BC_new =

a. [1%] Expected Error = |Expected – Observed| = |1 – 0| = 1
   [1% if correct, no partial points]
b. [2%] AC_new = AC + Alpha * Error * A_Output = -0.1 + 1 * 1 = 0.9
   [1% if only formula correct, 2% if final answer is correct]

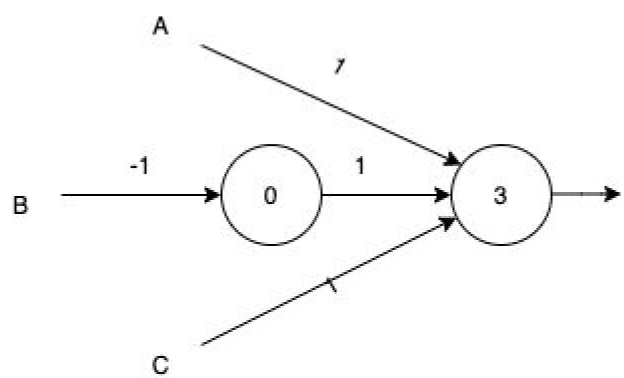
c. [2%] BC_new = BC + Alpha * Error * B_Output = 0.5 + 1 * 0 = 0.5
[1% if only formula correct, 2% if final answer is correct]

2.[6%] You are building a set of perceptrons to mimic logic circuits. For each perceptron assume it's firing is based on the threshold rule i.e.

$$f(x) = \begin{cases} 1 & \text{if } x \geq \theta \\ 0 & \text{otherwise} \end{cases}$$

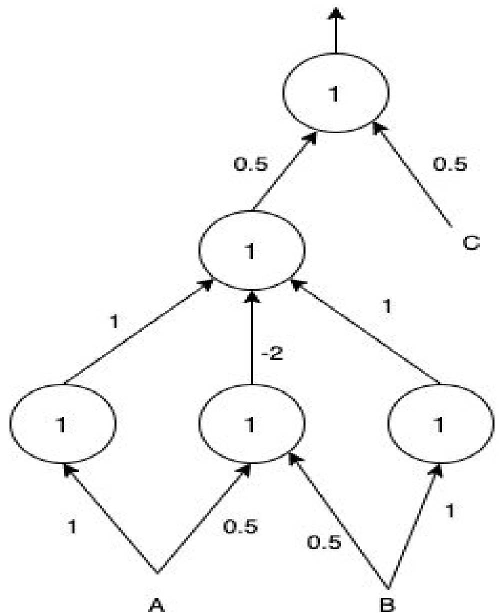
Assume all inputs have value either 0 or 1. Show how perceptrons can be used to mimic the following logic sentences. For each expression, (a) draw the perceptrons, (b) show the weights (c) write the threshold.

a[3%]. A ⋀ ¬B ⋀ C



- ¬B: 1.5 points, including weight and threshold
- The remaining: 1.5 points, assume ¬B is designed correctly
- If ¬B is not designed separately (has a single output itself), then grade the graph as a whole for 3 points.
- The grading will be comparing the outputs with the truth table, any failed case means the design is wrong and leads to 0 points on the corresponding part mentioned above. No more partial points.

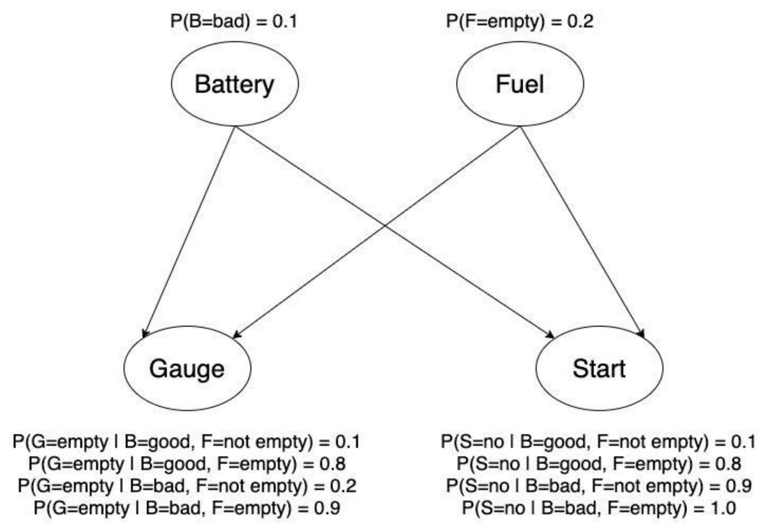
b[3%] (A ⊕ B) ⋀ C



- A ⊕ B: 1.5 points, including perceptrons, weights and thresholds
- The remaining: 1.5 points, assume A ⊕ B is designed correctly
- If A ⊕ B is not designed separately (has a single output itself), then grade the graph as a whole for 3 points.
- The grading will be comparing the outputs with the truth table, any failed case means the design is wrong and leads to 0 points on the corresponding part mentioned above. No more partial points

[15%] Bayesian Networks

Consider the following Bayesian Network with random variables
$Battery \in \{good, bad\}$, $Fuel, Gauge \in \{empty, not\ empty\}$, $Start \in \{yes, no\}$:



1.a. [4%] Find $P(B = good,\ F = not\ empty,\ G = empty,\ S = no)$ [For full credit, state the probability expression for the asked joint probability using the probabilities given in the Bayesian Network and show your work clearly].

2% for correct expression. 2% for correct solution.

a). $P(B = good,\ F = not\ empty,\ G = empty,\ S = no)$.

$= P(B = good) . P(F = not\ empty) . P(G = empty | B = g, F = ne) . P(S = no | B = g, F = ne)$.

$= 0.9 \times 0.8 \times 0.1 \times 0.1 = 0.0072$.

1.b [8%] Find $P(S|B = bad) = [P(S = yes|B = bad),\ P(S = no|B = bad)]$ **using enumeration** [For full credit, state the probability expression for the asked distribution using the probabilities given in the Bayesian Network and show your work clearly].

2% for correct expression. 4% for correctly computing . 2% for correct normalization.

b). $P(S \mid B=\text{bad})$

$$= \alpha \sum_f P(B=\text{bad}).P(F=f).P(S \mid B=\text{bad}, F=f).$$

$$= \alpha \begin{bmatrix} 0.1 \times 0.2 \times 0 + 0.1 \times 0.8 \times 0.1 \\ 0.1 \times 0.2 \times 1 + 0.1 \times 0.8 \times 0.9 \end{bmatrix}$$

$$= \alpha \begin{bmatrix} 0.008 \\ 0.02 + 0.072 \end{bmatrix} = \alpha \begin{bmatrix} 0.008 \\ 0.092 \end{bmatrix} = \begin{bmatrix} 0.08 \\ 0.92 \end{bmatrix}$$

$$\left(\because \alpha = \frac{1}{0.008 + 0.092} = \frac{1}{0.1} = 10\right).$$

1.c [3%] Consider the statements below. For each statement, mark *True* if it always holds in the context of the above Bayesian Network and *False* otherwise.

- $P(B,F) = P(B)P(F)$
- $P(B|F,G) = P(B|G)$
- $P(S,G|B,F) = P(S|B,F)P(G|B,F)$

1% for each correct response.

True
False
True

[18%] Probability Theory

Q1[9%] Mangoes are of only three mutually exclusive varieties i.e [Alphonso, Amrapali, Angie] (there are no other varieties). The color of the Mangoes is either 'Yellow' or 'Red' (there are no other colors). The taste of the mango is either Sweet or Sour (there are no other tastes). Let's Consider 'variety', 'taste' and 'color' as the random variables.

Also, you are given following probabilities :-

P(color = 'Yellow')
P(taste = Sweet)

P(taste = Sweet, ‘variety’ = Alphonso | color = ‘Yellow’)
P(taste = Sweet, ‘variety’ = Amrapali | color = ‘Yellow’)
P(taste = Sweet, ‘variety’ = Angie | color = ‘Yellow’)

In reference to the given “Mango” details above please answer the following (Please consider the probabilities given only in context of the above details for mangoes):-

A. **P(taste = Sweet | color = ‘Yellow’)**
B. **P(color = ‘Yellow’ | taste = Sweet)**
C. **P(‘variety’ = Alphonso) + P(‘variety’ = Amrapali) + P(‘variety’ = Angie)**

1) P(taste = Sweet | color = ‘Yellow’) = P(taste = Sweet, ‘variety’ = Alphonso | color = ‘Yellow’) + P(taste = Sweet, ‘variety’ = Amrapali | color = ‘Yellow’) + P(taste = Sweet, ‘variety’ = Angie | color = ‘Yellow’)
2) P(color = ‘Yellow’ | taste = Sweet) = P(taste = Sweet | color = ‘Yellow’) P(color = ‘Yellow’) / P(taste = Sweet)
~~3)~~ P(‘variety’ = Alphonso) + P(‘variety’ = Amrapali) + P(‘variety’ = Angie) = 1

3% Each

Q2[9%] Given the following joint distribution, answer the following questions:-

	cloudy		~cloudy	
	windy	~windy	windy	~windy
rain	0.108	0.012	0.072	0.008
~rain	0.016	0.064	0.144	0.576

Calculate :-

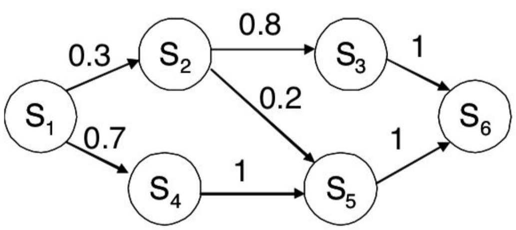
1) **P(~rain | cloudy)**
2) **P(cloudy or windy)**
3) **P(~windy)**

1) P(~rain | cloudy) = P(~rain and cloudy) / P(cloudy) = 0.4
2) P(cloudy or windy) = 0.416
3) P(~windy) = 0.66

3% Each

[12%] HMM

Consider the Hidden Markov Model defined by the transition and observation probabilities in the table below. The hidden variables are X_1, X_2, X_3, X_4, and the observations are O_1, O_2, O_3, O_4, i.e., the hidden and observed sequences have length 4. The hidden variables X_i can take one of six values, $\{s_1, \ldots, s_6\}$. The observations O_i are in {a, b, c, d}.
You know that $P(X_1 = s_1) = 1$, i.e., the hidden sequence starts in state s_1. The transition probabilities are as represented in the following transition diagram:



For example, $P(X_{t+1} = s_3 \mid X_t = s_2) = 0.8$, and $P(X_{t+1} = s_4 \mid X_t = s_1) = 0.7$. Edges which are missing correspond to transition probabilities of zero.
The observation probabilities are as follows:

	a	b	c	d
s_1	0.5	0.3	0	0.2
s_2	0.1	0.3	0.5	0.1
s_3	0.2	0.3	0.4	0.1
s_4	0.3	0.2	0.3	0.2
s_5	0	0.3	0.2	0.6
s_6	0.4	0.4	0.1	0.1

For example, $P(O_t = c \mid X_t = s_2) = 0.5$, and $P(O_t = a \mid X_t = s_4) = 0.3$. We will use the following shorthanded notation where we for example write $P(O = abca, X_2 = s_1, X_4 = s_2)$ instead of $P(O_1 = a, O_2 = b, O_3 = c, O_4 = a, X_2 = s_1, X_4 = s_2)$.

For each of the items below, insert <, > or = into the brackets between the left and the right expression.

1. $P(O=abca, X_1 = s_1, X_2 = s_2)$ (   ) $P(O=abca \mid X_1 = s_1, X_2 = s_2)$

2. $P(O=abca, X_1 = s_1, X_4 = s_6)$ (   ) $P(O=abca \mid X_1 = s_1, X_4 = s_6)$

3. $P(O=acdb, X_2 = s_2, X_3 = s_3)$ (   ) $P(O=acdb, X_2 = s_4, X_3 = s_5)$

4. $P(O=acdb)$ (   ) $P(O=acdb \mid X_2 = s_4, X_3 = s_5)$

Solutions:
3% per correct answer

1. "<". The right hand side is the left hand side divided by $P(X_1 = s_1, X_2 = s_2)$ which is less than 1.
2. "=". Here, $P(X_1 = s_1, X_4 = s_6) = 1$, hence we have equality.

3. "<". We work out $P(O = acdb, X_2 = s_2, X_3 = s_3) = C\, P(X_2 = s_2 \mid X_1 = s_1), P(X_3 = s_3 \mid X_2 = s_2) P(O_2 = c \mid S_2 = s_2)\, P(O_3 = d \mid S_3 = s_3) = C \cdot .3 \cdot .8 \cdot .5 \cdot .1$ which is less than $P(O = acdb, X_2 = s_4, X_3 = s_5) = CP(X_2 = s_4 \mid X_1 = s_1), P(X_3 = s_5 \mid X_2 = s_4) P(O_2 = c \mid S_2 = s_4) P(O_3 = d \mid S_3 = s_5) = C \cdot .7 \cdot 1 \cdot .3 \cdot .5$ for some constant C.
4. "<". We work out the probabilities $P(O = acdb, X_2 = s_2, X_3 = s_3)$, $P(O = acdb, X_2 = s_2, X_3 = s_5)$, $P(O = acdb, X_2 = s_4, X_3 = s_5)$ and sum them to get $P(O = acdb)$. To get $P(O = acdb \mid S_2 = s_4, S_3 = s_5)$, we can divide $P(O = acdb, X_2 = s_4, X_3 = s_5)$ by $P(X_2 = s_4, X_3 = s_5) = .7$.

[10%] Naive Bayes

Assume we have a data set with three binary independent input attributes, A, B, C, and one binary outcome attribute Y. The three input attributes, A, B, C take values in the set { 0, 1 } while the Y attribute takes values in the set { 0, 1 } .

A	B	C	Y
0	1	1	0
1	1	0	0
1	0	1	1
1	1	1	1
0	1	1	0
0	0	0	0
0	1	1	1
1	0	1	1
0	1	0	0
1	1	1	0

How would you classify the record S = (A = 0, B = 0, C = 1) based on the data given in the table above? Write down both $P(Y = 0 \mid A = 0, B = 0, C = 1)$ and $P(Y = 1 \mid A = 0, B = 0, C = 1)$, and explain if Y should be 0 or 1.

Rubric:

$P(Y = 0 \mid A = 0, B = 0, C = 1) = P(A = 0, B = 0, C = 1 \mid Y = 0)\ P(Y = 0) / P(A = 0, B = 0, C = 1)$

$P(Y = 1 \mid A = 0, B = 0, C = 1) = P(A = 0, B = 0, C = 1\ Y = 1)P(Y = 1) / P(A = 0, B = 0, C = 1)$

$P(A = 0 \mid Y = 0)P(B = 0 \mid Y = 0)P(C = 1 \mid Y = 0)P(Y = 0) = 4/6 * 1/6 * 3/6 * 6/10 = 1/30$. (3 points)

4/6 : 0.5 point, 1/6: 0.5 point, 3/6: 0.5 point, 6/10: 0.5 point

finally 1/30: 1 point

$P(A = 0 \mid Y = 1)P(B = 0 \mid Y = 1)P(C = 1 \mid Y = 1)P(Y = 1) = ¼ * 2/4 * 4/4 * 4/10 = 1/20$ (3 points)

1/4 : 0.5 point, 2/4: 0.5 point, 4/4: 0.5 point, 4/10: 0.5 point

finally 1/20: 1 point

$P(A = 0 \mid Y = 0)P(B = 0 \mid Y = 0)P(C = 1 \mid Y = 0)P(Y = 0) < P(A = 0 \mid Y = 1)P(B = 0 \mid Y = 1)P(C = 1 \mid Y = 1)P(Y = 1)$ (1 points) (optional)

$\therefore Y = 1$. (4 points)