ELG5255- Applied Machine Learning
Professor: Dr. Hitham Jleed
Student name :( )

01/11/2023
Midterm Exam (Fall 2023)
Student ID: ( )

This exam is 120 minutes long.
There should be four numbered pages in this exam.
This exam is closed book. It contains 4 questions.
Questions that ask you to “briefly explain" require (1-3 sentence) explanations.
Calculators are allowed, but no laptops or phones.

Question 1. Short Answers

a. Give an example of a loss function for classification problems.

b. Explain the trade-off between bias and variance.

c. How can increasing the complexity of a model affect its bias and variance?

d. Suppose you are given an EM algorithm that finds maximum likelihood estimates for a model with latent variables. You are asked to modify the algorithm so that it finds MAP estimates instead. Which step or steps do you need to modify?

1. Expectation 2. Maximization 3. No modification necessary 4. Both

Question 2:

(a) K-Means Clustering

- (True or False): K-Means is to converge (i.e., terminate).   (   )
- (True or False): A good way to pick the number of clusters. k, used for k-Means is to try multiple values of k and choose the value that minimizes the distortion measure. (   )
- (True/False) Running k-means with a larger value of k always enables a lower possible final objective value than running k-means with smaller k. (   )

(b) Discuss the importance of initialization in the k-means algorithm. How can poor initialization affect the results?

(c). Hierarchical Clustering.

Given three clusters. A, B and C. containing a total of six points, each point is defined by integer value in one dimension, A {O, 2, 6}, B {3, 9} and C {11}, which two dusters will be merged at the next iteration of Hierarchical Agglomerative Clustering (HAC) when Euclidean distance,

1. Single linkage?

   i. Merge A and B
   ii. Merge A and C
   iii. Merge B and C

2. Complete linkage?

   i. Merge A and B
   ii. Merge A and C
   iii. Merge B and C

Question 3:

(a). Two litters of a particular rodent species have been born, one with two brown-haired and one gray-haired (litter 1), and the other with three brown-haired and two gray haired (litter 2). We select a litter at random and then select an offspring at random from the selected litter. Find.

- The probability that the animal chosen is brown-haired $\equiv P(Brown\ hair)$
- Given that a brown-haired offspring was selected, the probability that the sampling was from litter 1 $\equiv P(Little1|Brown\ hair)$

(b). Below is the probability destiny function for the Rayleigh distribution,

$$f(x;\theta) = \frac{x}{\theta^2} \exp\left(\frac{-x^2}{2\theta^2}\right)$$

Suppose you gathered data consisting of $x_1, x_2, \dots, x_n$ which deems are iid Rayleigh(θ) random variables.

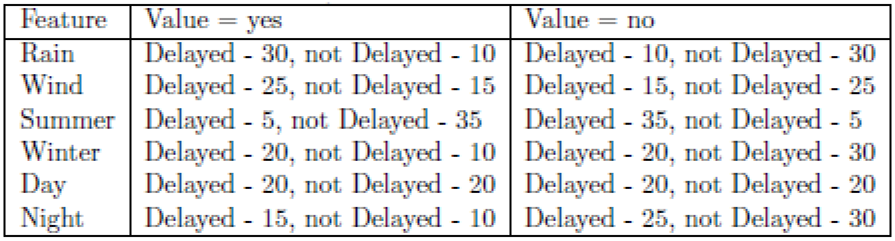
$l(\theta|X) \equiv$

$\mathcal{L}(\theta|X) \equiv$

The maximum likelihood estimate of Rayleigh's parameter. $\hat{\theta}_{MLE} \equiv$

Question 4:

(a). Our goal is to construct a decision tree classifier for predicting flight delays. We have collected data for a few months and a summary of the data is provided in the following Table.



Feature	Value = yes	Value = no
Rain	Delayed - 30, not Delayed - 10	Delayed - 10, not Delayed - 30
Wind	Delayed - 25, not Delayed - 15	Delayed - 15, not Delayed - 25
Summer	Delayed - 5, not Delayed - 35	Delayed - 35, not Delayed - 5
Winter	Delayed - 20, not Delayed - 10	Delayed - 20, not Delayed - 30
Day	Delayed - 20, not Delayed - 20	Delayed - 20, not Delayed - 20
Night	Delayed - 15, not Delayed - 10	Delayed - 25, not Delayed - 30

i. Based on the table, which feature should be at the root of the decision tree (briefly explain, no need to provide exact values for information gain)?

ii. Based on the table, which feature should be on the second level (the level just beneath the root) of the decision tree (briefly explain, no need to provide exact values for information gain)?

(b). Given the following data of transactions at a shop, calculate the support and confidence values of milk —> bananas, bananas —> milk, milk —> chocolate, and chocolate —> milk.

Transaction	Items in basket
1	milk, bananas, chocolate
2	milk, chocolate
3	milk, bananas
4	chocolate
5	chocolate
6	milk, chocolate

milk —> bananas : Support = ( ) Confidence: ( ).

bananas —> milk : Support = ( ) Confidence: ( ).

milk —> chocolate : Support = ( ) Confidence: ( ).

chocolate —> milk : Support = ( ) Confidence: ( ).

ELG5255- Applied Machine Learning
Professor: Dr. Hitham Jleed

uOttawa

01/11/2023
Midterm Exam (Fall 2023)

General Formula Sheet

Training set contains N such examples, $X = \{x^t, r^t\}_{t=1}^N$. ML approximates r^t using the model $g(x^t|\theta)$.
Loss function, L(·), to compute the difference between r^t and $g(x^t|\theta)$.

The *approximation error*, $E(\theta|X) = \sum L(r^t, g(x^t|\theta))$.

Observed variables, x, The Bayes' rule is, $P(C|x) = P(C)P(x|C)/P(x)$

Association Rules

$$\text{Support}(X,Y) \equiv P(X,Y) = \frac{\#\{\text{customers who bought } X \text{ and } Y\}}{\#\{\text{customers}\}}$$

$$\text{Confidence}(X \rightarrow Y) \equiv P(Y|X) = \frac{P(X,Y)}{P(X)} = \frac{\#\{\text{customers who bought } X \text{ and } Y\}}{\#\{\text{customers who bought } X\}}$$

$$\text{Lift}(X \rightarrow Y) = \frac{P(X,Y)}{P(X)P(Y)} = \frac{P(Y|X)}{P(Y)}$$

LIKELIHOOD $$l(\theta|\mathcal{X}) \equiv p(\mathcal{X}|\theta) = \prod_{t=1}^{N} p(x^t|\theta)$$

LOG LIKELIHOOD $$\mathcal{L}(\theta|\mathcal{X}) \equiv \log l(\theta|\mathcal{X}) = \sum_{t=1}^{N} \log p(x^t|\theta)$$

$$\begin{aligned} \text{E-step} &: \quad \mathcal{Q}(\Phi|\Phi^l) = E[\mathcal{L}_c(\Phi|\mathcal{X},\mathcal{Z})|\mathcal{X},\Phi^l] \\ \text{M-step} &: \quad \Phi^{l+1} = \arg\max_{\Phi} \mathcal{Q}(\Phi|\Phi^l) \end{aligned}$$

ENTROPY $$\mathcal{I}_m = -\sum_{i=1}^{K} p_m^i \log_2 p_m^i$$

Gini Index, $G_m = \sum_{i=1}^K p_m^i(1 - p_m^i)$

Information Gain (IG)= entropy(parent-node) – [average entropy(children-nodes)]

Good Luck