University of Asia Pacific

Admit Card

Mid-Term Examination of Spring, 2021

Financial Clearance	PAID

Registration No : 18101064

Student Name : Md. Sohanuzzaman Soad

Program : Bachelor of Science in Computer Science and Engineering

SI.NO.	COURSE CODE	COURSE TITLE	CR.HR.	EXAM. SCHEDULE
1	CSE 400	Project / Thesis	3.00	
2	CSE 401	Mathematics for computer Science	3.00	
3	CSE 403	Artificial Intelligence and Expert Systems	3.00	
4	CSE 404	Artificial Intelligence and Expert Systems Lab	1.50	
5	CSE 405	Operating Systems	3.00	
6	CSE 406	Operating Systems Lab	1.50	
7	CSE 407	ICTLaw, Policy and Ethics	2.00	
8	CSE 410	Software Development	1.50	
9	CSE 427	Topics of Current Interest	3.00	

Total Credit: 21.50

1. Examinees are not allowed to enter the examination hall after 30 minutes of commencement of examination for mid semester examinations and 60 minutes for semester final examinations.

2. No examinees shall be allowed to submit their answer scripts before 50% of the allocated time of examination has elapsed.

3. No examinees would be allowed to go to washroom within the first 60 minutes of final examinations.

4. No student will be allowed to carry any books, bags, extra paper or cellular phone or objectionable items/incriminating paper in the examination hall.
Violators will be subjects to disciplinary action.

This is a system generated Admit Card. No signature is required.

UNIVERSITY OF ASIA PACIFIC

Department of Computer Science & Engineering

Mid-Term Examination
Spring-2021

Student Name	:	Md. Sohanuzzaman Soad
Student ID	:	18101064
Section	:	B
Year	:	4^{th}
Semester	:	1^{st}
Course Code	:	CSE 427
Course Title	:	Topics of Current Interest
Date	:	18-Septempber-2021

18101064

Ans to the Que. No: 1(a)

Data type are a way of classification that specifies which type of value a variable can store and what type of mathematical operations and relation or logical operation can be applied to the variable without any error. In machine learning, it is very important to know exact datatype of independent and dependent variable.

Unit of observation: unit of observation is the smallest entity with measured properties of ~~intest~~ interest for a study.

18101064

Examples:

A person, A time point,

A measurement.

Dataset that store the units of observation and their properties can be imagined as collection of data consisting of the following:

i) Example: An "Example" is an instance of the unit of observation for which properties have been recorded.

ii) Features: A "Feature" is a recorded property or a characteristic of examples.

18101064

Example of "Example" and "Features" :

Cancer detection : Consider the problem of developing an algorithm for detecting cancer. For this case study we note the following :

(a) The units of observation are the patients

(b) The examples are member of a sample of cancer patients.

(c) The following attributes of the patients may be choosen as the features:

(i) Gender (ii) Age (iii) Blood presure

(iv) the findings of the pathology report after a biopsy.

Different forms of Data:

1) Numeric Data: If a feature represents a characteristic measured in numbers, it is called numeric ~~nominal~~ feature

2) Categorical or Nominal: A categorical feature is an attribute that take on one of a limited, and usually fixed number of possible values on the basis of some qualitative property.

18101064

5

3 ordinal Data: This denotes a nominal variable with categories falling in an ordered list.

Examples:

Features

Year	model	price	transmission
2011	sel	21000	Auto
2012	sel	22000	Manual

Examples

In the given data the features "Year", "Price" are numeric and "model" "transmisson" are categorical.

Ans to the Que. No: 1(b)

Noise: Noise is any unwanted anomaly in the data. Noise may arise due to several factors:

1) There may be imprecision in recording the input ~~numbers~~ attributes, which may shift the data points in the input space.

2) There may be errors in labeling the data points, which may relabel positive instances as negative and vice versa.

③ There may be additional Attributes, which we have not into account. that affect the ~~b~~ label of an instance

Effect of Noise:

Noise ~~do~~ distorts data, when there is noise in data, learning problem may not produce accurate results. Also, simple hypotheses may not be sufficient to explain the data and so complicated hypotheses may have to formulated. This leads to use the additional computing resources and the needless wastage of such resources.

Ans to the Que. No: 2(a)

Example: Consider a supermarket chain.

The management of the chain is interested in knowing whether there are any patterns in the purchases of product by customer like the following:

"If a customer buy onions and potatoes together, then he/she is likely to also buy hamburger"

From the standpoint of customer behaviour, this defines an association between the set of products {onion, potato} and

18101064

{burger}. This association is represented in the form of a rule as follows:

$$\{onion, potato\} \Rightarrow \{burger\}$$

The probablity: (conditional probablity)

$$P(\{onion, potato\} \mid \{burger\})$$

If this conditional ~~propa~~ probablity is 0.8, then the rule may be stated as;

"80% of customers who buy onion and potato also buy burger"

18101064

How association rules are made use of

consider an association rule of the

form :

$$X \Rightarrow Y$$

That is, if people ~~buy~~ X they ~~tee~~ also buy Y.

18101064

Ans to the Que. No: 2(b)

$$X = (x_1, x_2, \ldots, x_n)$$

$$P(C_1|X), P(C_2|X), \ldots, P(C_p|X) \quad \text{(i)}$$

using Bayes' theorem, we have,

$$P(C_K|X) = \frac{P(X|C_K)\,P(C_K)}{P(X)} \quad \text{(ii)}$$

Hence we have,

$$P(X|C_K) = P((x_1, x_2, \ldots x_n)|C_K)$$

$$= P(x_1|C_K)\,P(x_2|C_K) \ldots P(x_n|C_K)$$

using eqn (ii) we get,

$$P(C_K|X) = \frac{P(x_1|C_K)\,P(x_2|C_K) \ldots P(x_n|C_K)\,P(C_K)}{P(X)}$$

Since the denominator $P(X)$ is indepen-dent of the class labels, we have

$$P(c_K|x) \propto P(x_1|c_K)\,P(x_2|c_K) \cdots P(x_n|c_K)\,P(c_K)$$

So, it is enough to find the maximum among the following values

$$P(x_1|c_K)\,P(x_2|c_K) \_\_ P(x_n|c_K)\,P(c_K) \; [K=1, \ldots P]$$

18101064

Ans to the Que. No: 3(a)

~~char~~ classify (red, SUV, domestic) using naive Bayes' Algorithm:

Frequency table:

	Yes	No
Red	3	2
Yellow	2	3

	Yes	No
Sports	4	2
SUV	1	3

	Yes	No
Domestic	2	3
Imported	3	2

18101064

likelihood Table:

	Yes	No
red	3/5	2/5
Yellow	2/5	3/5

	Yes	No
Sports	4/5	2/5
SUV	1/5	3/5

	Yes	No
Domestic	2/5	3/5
Imported	3/5	2/5

Now calculate the following numbers:

$P(Yes \mid X) = P(red \mid Yes)\, P(SUV \mid Yes)$
$\quad P(domestic \mid Yes)\, P(Yes)$
$= (3/5)(1/5)(2/5)(5/10)$
$= 0.024$

$$P(No \mid X) = P(Red \mid No)\, P(SUV \mid No)\, P(Domestic \mid No)\, P(No)$$

$$= (2/5)(3/5)(3/5)(5/10)$$

$$= 0.072$$

The maximum value is 0.072

Therefore X belongs to class (stolen = No)

Since 0.144 > 0.048 which means given features RED, SUV, Domestic our example get classified as "No" the car is not stolen.

18101064

Ans to the Que. No 3 (b)

In a ~~Bernoli~~ Bernoulli distribution there are two outcomes. An event occurs or it does not.

Let x is:

$$f(x|p) = p^{x} (1-p)^{1-x} \quad , x = 0,1$$

~~estipa~~ estimation of p:

$$L(p) = \log f(x_1|p) + \_\_ + \log f(x_n|p)$$

$$= \log p^{x_1} (1-p)^{1-x} + \_\_\_\_ + \log p^{x_n} (1-p)^{1-x_n}$$

$$= [x_1 \log p + (1-x_1)\log(1-p)] + \_\_ + [x_n \log p + (1-x_n)\log(1-p)]$$

p that maximize $L(p)$ we set up the equation

$$\frac{dL}{dp} = 0$$

that is.

$$\left[\frac{x_1}{p} - \frac{1-x_1}{1-p}\right] + \dots + \left[\frac{x_n}{p} - \frac{1-x_n}{1-p}\right] = 0$$

solve the eqn we get,

$$p = \frac{1}{p}(x_1 + \dots + x_n)$$