

Financial Mathematics (FM-410)

Python Programming for Data Sciences

Department of Mathematics (Final Examination) University of Karachi

Lecturer: Engineer Syed Umaid Ahmed

Time Allowed: 3 hours

Note: Attempt any five questions. All carry equal marks

Total Marks: 50

Question 1:

Table I shows a confusion matrix for a classification problem. Calculate the following measures:

	Predicted +	Predicted -
True +	100	40
True -	60	300

- (a) What is Confusion Matrix? Why we used it for model evaluation?
- (c) Comment on the quality or performance of the model
- (d) Name the Python Library to find confusion matrix of a dataset.

Question 2:

Draw a decision tree by selecting the root node. Make the necessary calculation of (a) Information Gain and (b) Entropy.

Weather	Temperature	Humidity	Wind	Golf Play
Fine	Hot	High	None	No
Fine	Hot	High	Few	No
Cloud	Hot	High	None	Yes
Rain	Warm	High	None	Yes
Rain	Cold	Medium	None	Yes
Rain	Cold	Medium	Few	No
Cloud	Cold	Medium	Few	Yes
Fine	Warm	High	None	No
Fine	Cold	Medium	None	Yes
Rain	Warm	Medium	None	Yes
Fine	Warm	Medium	Few	Yes
Cloud	Warm	High	Few	Yes
Cloud	Hot	Medium	None	Yes
Rain	Warm	High	Few	No

**Question 3:**

- (a) What is the difference between a box plot and a histogram? Explain the use of blox plot with reference to a problem of Titanic Dataset.
- (b) Consider a dataset having two variables weight (kg) and height (cm) and each point is classified as normal or under-weight.

Weight (x2)	Height (y2)	Class
51	167	Underweight
62	182	Normal
69	176	Normal
64	173	Normal
65	172	Normal
56	174	Underweight
58	169	Normal
57	173	Normal
55	170	Normal

Suppose I want to classify the data point having weight as 57 kg and height 170 cm. By using the mathematical calculation and Euclidean distance formula classify its nearest neighbors. Also show your calculation with the help of datapoints on a graph.

**Question 4:**

- (a) What is the difference between Linear Regression and Logistic Regression? Discuss both of them with reference to real world examples.
- (b) Here is dataset of hiring shared by the HR of the company.

Experience	Test Score (10)	Interview (10)	Salary \$
	8	9	50000
	8	6	45000
5	6	7	60000
2	10	10	65000
7	9	6	70000
3	7	10	62000
10		7	72000
11	7	8	80000

- i. Share the complete strategy to work on this dataset.
- ii. Suggest the best Machine Learning Model for finding the salary of person having 5,6,5 as datapoint in first three columns

**Question 5:**

- (a) Why do we need data sciences? Give an example of a single problem you have solved in the course at University of Karachi.
- (b) Differentiate between descriptive and inferential statistics with reference to a practical example. Also, name the measures of central tendency.
- (c) What is meant by data cleaning and pre-processing ?
- (d) Name five important libraries you've used in real world data analysis projects.

**Question 6:**

- (a) What is Normal Distribution? Explain it with the help of figure & its connection to the standard deviation.
- (b) Briefly explain any two types of Probability?
- (c) Here is the dataset given for the training study undergone at University of Karachi. 60 candidates without training and 40 candidates are with institution training. Examine the dataset and answer the following questions.

	Results	Training		Total
		Without Training	With Training	
Salary Packages Obtained by Students	Very Poor Package	5	0	5
	Poor Package	10	0	10
	Average Package	40	10	50
	Good Package	5	30	35
	Excellent Package	0	5	5
	Total	60	45	105

- i. Find the Probability that a candidate has gone University Training. Name this type of Probability.
- ii. Find the Probability that a candidate has gone University's Training and a Good Package. Name this type of Probability.
- iii. Find the Probability that a candidate has good package and not undergone University's Training. Name this type as well.

***Question 7:***

- (a) At a party there were 66 handshakes. Everyone shook hands with everybody else. How many people were present at the party?
- (b) Explain Null Hypothesis and Alternative Hypothesis?
- (c) What is Baye's Theorem and Naïve Baye's Technique in Machine Learning. Discuss real world example of these powerful Machine Learning Techniques.

**We wish you best of luck for your future endeavors**