९/41

3012N1147

Candidate's Seat No:____

MBA in BI Sem.-3 Examination
BI-E107
FML
December-2025

Time : 1.30 Hours] **[Max.Marks : 70**

Instructions :(1) This paper contains **Thirty Five** questions.
(2) Each Question is of 2 Marks.
(3) Each Question is of multiple choices.
(4) All questions are compulsory.

No.	QUESTIONS	Marks
Q.1	Which statement best distinguishes AI from Machine Learning?	**02**
	A. AI focuses solely on statistical learning, while ML includes symbolic reasoning	B. ML is a subset of AI focused on learning patterns from data
	C. AI deals only with rule-based logic, while ML deals only with neural networks	D. ML requires human-defined rules, whereas AI learns rules automatically
Q.2	Which of the following best characterizes the goal of supervised learning?	**02**
	A. Discovering hidden structures without output labels	B. Learning a function that maps inputs to outputs using labeled data
	C. Minimizing reinforcement signals from an environment	D. Generating synthetic data from a latent space
Q.3	Which scenario most strongly indicates a classification problem rather than regression?	**02**
	A. Predicting tomorrow's temperature	B. Determining the sentiment of a customer review
	C. Estimating the price of a used car	D. Forecasting a continuous time series
Q.4	What is the primary role of a loss function in machine learning?	**02**
	A. Constraining model memory usage	B. Determining the model's architecture

P.T.O

N1147 -2

C. Quantifying the discrepancy between predictions and true values

D. Selecting the optimal learning rate

Q.5 What is the primary goal of simple linear regression? **02**

A. To classify data into categories

B. To model the relationship between a single independent variable and a dependent variable as a straight line

C. To predict multiple outcomes simultaneously

D. To cluster similar data points

Q.6 In the equation $y = mx + b$, what does *'m'* represent? **02**

A. The y-intercept

B. The slope of the line, indicating the change in y per unit change in x

C. The error term

D. The correlation coefficient

Q.7 Which NumPy function is used in the notebook to fit a linear model to the data? **02**

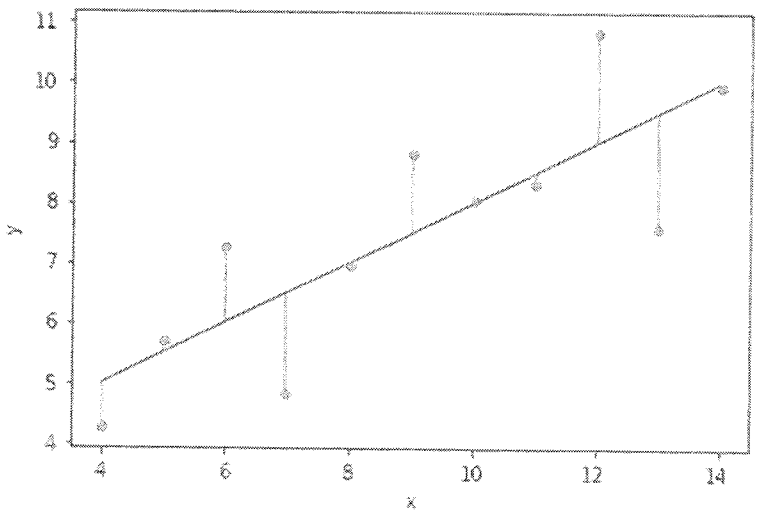
A. np.corrcoef()

B. np.polyfit()

C. np.linspace()

D. np.histogram()

Q.8 What is the purpose of the red line plotted over the scatterplot in the notebook? **02**



A. To show the data points more clearly

B. To represent the predicted sales based on a linear fit to total_spend

C. To highlight outliers in the data

D. To indicate the correlation coefficien

Q.9 What is a key difference between simple linear regression and multiple linear regression? **02**

A. Simple LR uses categorical variables; multiple LR uses continuous ones

B. Simple LR models one independent variable; LR multiple models two or more

C. Simple LR requires normalization; multiple LR does not

D. Simple LR predicts sales; LR multiple predicts budgets

Q.10 Given the dataset characteristics from df.describe() (age: mean=51.609, std=11.287; physical_score: mean=32.760, std=8.169; test_result: mean=0.6), and knowing the dataset has 5000 samples with 60% positive class (test_result=1), which of the following preprocessing steps is MOST critical before fitting a LogisticRegression model, and why? **02**

A. One-hot encoding categorical variables, because age and physical_score are ordinal and could introduce bias in the sigmoid function.

B. Feature scaling using StandardScaler, as logistic regression's gradient descent optimization converges faster and avoids feature dominance due to differing scales (e.g., age up to 90 vs. physical_score up to 50).

C. Log transformation on test_result, since the target is imbalanced and log-loss assumes Gaussian distribution.

D. Polynomial feature generation up to degree 2, to capture non-linear interactions between age and physical_score without assuming linearity in logit space.

Q.11 The value_counts() shows 3000 positives (pass) and 2000 negatives (fail), yielding a baseline accuracy of 60% by always predicting 'pass'. After train-test split (test_size=0.3, random_state=42) and fitting the model, suppose the confusion matrix on test data is [[500, 200], [300, 1000]] (TN=500, FP=200, FN=300, TP=1000). What is the model's precision for the positive class, and how does it compare to accuracy in this imbalanced scenario? **02**

A. Precision = 1000/(1000+200) = 83.3%; superior to accuracy (75%) because it focuses on false positives, critical for medical tests where over-

B. Precision = 1000/(1000+300) = 76.9%; inferior to accuracy (78.3%) due to class imbalance inflating accuracy via majority class.

P.T.O

diagnosing hearing ability (FP) is costly.

C. Precision = (1000+500)/(1500+500) = 83.3%; equivalent to accuracy, as the split preserves the 60:40 ratio perfectly.

D. Precision = 1000/1500 = 66.7%; better than accuracy (78.3%) since recall is prioritized in binary classification.

Q.12 The Iris dataset in the notebook has 150 samples equally distributed across three species (50 each). After a train_test_split with test_size=0.3 and random_state=101, approximately how many samples are in the test set for each class, assuming stratified splitting? **02**

A. 15 samples per class (total 45).

B. 20 samples per class (total 60)

C. 25 samples per class (total 75)

D. 30 samples per class (total 90)

Q.13 The notebook fits StandardScaler on X_train and transforms both train and test sets. Given Iris features (e.g., sepal_length mean≈5.84, std≈0.83; petal_length mean≈3.76, std≈1.77), why is scaling essential for logistic regression? **02**

A. It ensures the sigmoid/softmax outputs probabilities summing to 1 across classes.

B. It prevents features with larger scales (e.g., sepal_length) from dominating the linear predictor, improving coefficient interpretability and convergence.

C. It handles outliers in petal_width (min=0.1), which could otherwise cause numerical instability in OvR classifiers.

D. It converts features to binary (0/1) for compatibility with multi_class='ovr'.

Q.14 In the classification_report, metrics like precision, recall, and F1 are computed per class and averaged (macro/micro). For a well-separated dataset like Iris, the macro F1-score is typically close to 1.0. What does macro-averaging emphasize in multi-class evaluation? **02**

A. Performance on the majority class, weighting by sample frequency.

B. Equal weight to each class's F1, highlighting issues in minority or harder-to-classify classes (e.g., versicolor vs. virginica overlap)

C. Harmonic mean across all predictions, ignoring class labels.

D. Threshold-based scoring using predict_proba outputs.

Q.15 What is the primary objective when selecting a split point in a decision tree algorithm during the training process?. **02**

A. To maximize the depth of the tree for better granularity

B. To minimize the weighted impurity (e.g., Gini or entropy) across child nodes, ensuring homogeneous subsets

C. To balance the number of samples in each child node equally

D. To prioritize categorical features over numerical ones for faster computation

Q.16 In a binary decision tree for classification, which of the following best describes the role of the Gini impurity index when evaluating potential splits? **02**

A. It measures the probability of misclassifying a randomly chosen sample if labeled according to the class distribution in a node (range: 0 to 0.5)

B. It calculates the information gain relative to the parent node, always favoring splits with higher entropy

C. It quantifies the variance reduction for regression tasks only, ignoring class labels

D. It penalizes deep trees by incorporating the tree's height into the split score

Q.17 Which of the following is a key advantage of decision trees over linear models like logistic regression in handling complex datasets? **02**

A. They assume linear separability and require feature scaling for convergence

B. They are non-parametric, capturing non-linear interactions and hierarchies without assuming data distribution

C. They provide probabilistic outputs via softmax, ideal for multi-class problems

D. They inherently handle multicollinearity by orthogonalizing features during splits

Q.18 What is the core principle behind Random Forests as an ensemble method? **02**

A. Averaging predictions from neural networks to reduce variance

B. Building multiple decision trees on bootstrapped data subsets and random feature selections, then aggregating outputs via voting (classification) or averaging (regression)

C. Stacking logistic regression models with boosting to minimize bias

D. Pruning individual SVMs to create a forest of support vectors

N 11417-6

Q.19 In Random Forests, what role does the bootstrap sampling (bagging) play in the training process? **02**

A. It ensures all features are used in every tree for computational efficiency

B. It creates diverse training sets by sampling with replacement (typically 63% unique samples per tree), reducing correlation among trees

C. It splits the data deterministically into training and validation sets without randomness

D. It applies only to regression tasks, using median instead of mean for outliers

Q.20 In a Random Forest for classification with 100 trees, how is the final prediction made for a new sample? **02**

A. By selecting the tree with the highest confidence score

B. Via majority voting across all tree predictions, with optional weighting by tree accuracy

C. By averaging the probability outputs and applying a softmax function

D. Through a greedy merge of the top-k tree decisions

Q.21 The initial RandomForestClassifier is fit with n_estimators=10 and random_state=101. What does n_estimators represent in RF, and why start with a small value like 10? **02**

A. Number of features per split; small for faster Gini computation

B. Number of decision trees in the forest; small for quick baseline before tuning

C. Bootstrap sample size; small to reduce OOB error variance

D. Maximum depth per tree; small to prevent early overfitting

Q.22 What is the primary goal of a Support Vector Machine (SVM) in binary classification? **02**

A. To minimize the sum of squared errors between predictions and actual labels

B. To find the hyperplane that maximizes the margin between the two classes while minimizing classification errors

C. To cluster data points into groups based on similarity

D. To reduce the dimensionality of features using principal components

Q.23 The notebook uses a linear kernel SVM. What is a key assumption of the linear kernel, and when is it most appropriate? **02**

A. Data is non-linear; appropriate for high-dimensional spaces

B. Classes are linearly separable; appropriate for simple, straight-line boundaries

C. Features are categorical; appropriate after one-hot encoding

D. Targets are multi-class; appropriate with one-vs-one strategy

Q.24 In SVM fundamentals, what are "support vectors," and why are they crucial? **02**

A. All data points used in training; they define the model's weights

B. The data points closest to the decision boundary; they determine the hyperplane's position

C. Outlier points removed during preprocessing; they prevent overfitting

D. Vectors in the kernel trick transformation; they enable multi-class extension

Q.25 In SVM, what is the difference between hard-margin and soft-margin classification? **02**

A. Hard-margin allows no errors; soft-margin penalizes but permits some via slack variables

B. Hard-margin uses RBF; soft-margin uses linear kernels

C. Hard-margin for regression; soft-margin for classification

D. Hard-margin multi-class; soft-margin binary only

Q.26 What is a key advantage of SVM over logistic regression for classification? **02**

A. SVM provides probabilistic outputs natively

B. SVM's margin maximization leads to better generalization in high-dimensional spaces, robust to outliers via support vectors

C. SVM handles missing values without imputation

D. SVM is faster for large datasets without kernels

Q.27 What is the primary purpose of tokenization in NLP? **02**

A. To convert text into numerical vectors for machine learning models

B. To break down text into smaller units like words or sentences for further processing

C. To identify sentiment or emotional tone in the text

D. To remove duplicate words and normalize frequencies

P.T.O

N 11428

Q.28 In the notebook, NLTK's 'punkt' resource is downloaded. What does it specifically enable? **02**

A. Part-of-speech tagging for words

B. Pre-trained tokenization models for sentence and word boundaries

C. Lemmatization using WordNet

D. Named entity recognition across languages

Q.29 What are stopwords in NLP, and why are they needed to be removed ? **02**

A. High-frequency words like "the" or "is"; removed to focus on meaningful content words

B. Rare words with low frequency; removed to reduce computational noise

C. Proper nouns; removed for privacy in text processing

D. Emojis or special characters; removed for Unicode normalization

Q.30 What is the difference between stemming and lemmatization in NLP fundamentals? **02**

A. Stemming preserves context; lemmatization is rule-based

B. Stemming chops word endings (e.g., "running" → "run"); lemmatization uses vocabulary for base form (e.g., "better" → "good")

C. Stemming requires POS tags; lemmatization does not

D. Stemming is for sentences; lemmatization is for words

Q.31 The notebook downloads 'wordnet' from NLTK. What is WordNet primarily used for in NLP? **02**

A. A lexical database for synonym sets (synsets) and lemmatization

B. Tokenization of multilingual text

C. Training sentiment classifiers

D. Generating n-grams automatically

Q.32 In POS tagging, what does the 'averaged_perceptron_tagger' resource enable? **02**

A. Identifying noun phrases only

B. Assigning parts-of-speech labels (e.g., NN for noun) to tokens

C. Clustering similar tags across documents

D. Translating tags to different languages

Q.33 What is sentiment analysis in NLP fundamentals? **02**

A. Measuring text readability scores

B. Classifying text as positive, negative, or neutral based on polarity

N1147-9

C. Extracting keywords via TF-IDF

D. Translating text to multiple languages

Q.34 What is the primary goal of Principal Component Analysis (PCA)? **02**

A. To classify data points into predefined categories

B. To reduce the dimensionality of a dataset while preserving as much variance as possible

C. To impute missing values in a dataset

D. To normalize features to a zero mean and unit variance

Q.35 In PCA, what does a principal component represent? **02**

A. The mean value of a single feature across all samples

B. A linear combination of original features that captures the direction of maximum variance

C. The eigenvector corresponding to the smallest eigenvalue of the covariance matrix

D. A categorical grouping of similar data points