1: (MCQ) Topic: Precision and Recall | Level: Medium

Question:
A hospital builds a machine learning model to detect whether a patient has a rare disease. The disease is very dangerous but also very rare. After initial testing, the doctors notice that the model often predicts "no disease" even when a patient actually has it.

Which of the following should the hospital prioritize improving?

- (a) Precision 2
- (b) Recall
- (c) Accuracy
- (d) F1-score

Correct Answer: (b) Recall

Explanation:
In this case, **missing a patient who actually has the disease is very risky**.

- **Recall** measures how many of the actual positive cases (patients who have the disease) the model correctly identifies.
- A **low recall** means many real cases are being missed (false negatives).
  Thus, **improving recall** is crucial to ensure that as few diseased patients as possible are overlooked, even if it means a few healthy people are incorrectly flagged.

Correct Answer: (b)
Explanation: Linear regression assumes that the residuals (errors) follow a normal distribution. Other assumptions include linearity, homoscedasticity, and no multicollinearity.

2. (MCQ) Topic: Linear Regression | Level: Easy

Question: A real estate company develops a regression model to predict house prices based on square footage. The model equation is:

$$Price = 50000 + 150 * (SquareFootage)$$

What is the predicted price of a house with 2000 square feet?

- (a) 300,000
- (b) 350,000
- (c) 400,000

- (d) 450,000

Correct Answer: (b)
Explanation: Plugging in 2000 for Square Footage:

$$50000 + 150 * 2000 = 350000.$$

3. (MCQ) Topic: Linear Regression | Level: Medium

Question: A financial analyst fits a linear regression model to predict stock prices based on historical data. The model gives an R^2 value of 0.78. What does this imply?

- (a) 78% of the variance in stock prices is explained by the model.
- (b) The model is overfitted.
- (c) The model has a 78% accuracy rate.
- (d) The model's error is 22%.

Correct Answer: (a)
Explanation: R^2 represents the proportion of variance explained by the independent variables. A value of 0.78 indicates that 78% of the variance in stock prices is explained by the model.

4. (MCQ) Topic: Linear Regression | Level: Hard

Question: A model predicts the following house prices: {200,000, 250,000, 300,000}. The actual prices are {210,000, 240,000, 310,000}. Calculate the Mean Squared Error (MSE).

- (a) 100,000,000
- (b) 50,000,000
- (c) 200,000,000
- (d) 300,000,000

Correct Answer: (a)
Explanation:

$$MSE = \frac{(200000 - 210000)^2 + (250000 - 240000)^2 + (300000 - 310000)^2}{3} = 100,000,000.$$

5. (MCQ) Topic: Regularization | Level: Medium

Question: A startup is predicting customer churn using linear regression. The training error is very low, but the test error is high. What does this indicate?

- (a) The model is underfitting.
- (b) The model is overfitting.
- (c) The model has a low bias and low variance.
- (d) The model is using an inappropriate loss function.

Correct Answer: (b)
Explanation: Overfitting occurs when the model performs well on training data but poorly on test data. This suggests the model is too complex and captures noise instead of the true relationship.

6. (NAT) Topic: Gradient Descent | Level: Medium

Question:
A company is training a machine learning model using **gradient descent**. Given the update equation:

$$w' = w - \alpha \frac{dJ}{dw}$$

If:

- **Initial weight** $w = 12$
- **Learning rate** $\alpha = 0.05$
- **Gradient** $\frac{dJ}{dw} = 8$

What is the updated value of w'?

Correct Answer: 11.6
Explanation:

$$w' = 12 - (0.05 \times 8) = 12 - 0.4 = 11.6$$

The **new weight value is 11.6** after applying the update rule.

7. (MCQ) Topic: Classification | Level: Easy

Question: A telecom company wants to predict whether a customer will leave the service (churn) or stay.
Which of the following is the best type of machine learning model for this problem?

- (a) Regression
- (b) Classification
- (c) Clustering
- (d) Reinforcement Learning

Correct Answer: (b)
Explanation: The goal is to **classify** customers as **"churn" or "not churn"**, making this a classification problem.

Question 8:

A student analyzes stock market trends and uses PCA to extract patterns. They observe that one principal component explains 90% of the data variance. What should they conclude?

A. The dataset is too simple

B. One feature dominates all others

C. Most variability in the data is captured by this direction

D. PCA failed to work

Answer: C

Explanation: A principal component that explains a large proportion of the variance indicates that most of the data's variability is concentrated along one direction. This often means that the data can be effectively represented in a lower-dimensional space without significant loss of information. It's a sign that PCA worked well in finding a meaningful pattern in the dataset.

9. (MCQ) Topic: SVM | Level: Medium

Question: Aryan is training an SVM model for digit classification. He notices that increasing a certain parameter makes the model too rigid, while decreasing it allows some misclassifications but improves

generalization. Which parameter is he adjusting?

- (a) Learning rate
- (b) Regularization strength
- (c) Kernel size
- (d) Number of clusters

Correct Answer: (b)
Explanation: The parameter **C** in SVM controls **the trade-off** between maximizing the margin and minimizing the training error.

10. (MCQ) Topic: SVM | Level: Hard

Question: A financial institution is developing a fraud detection system to classify transactions as fraudulent or legitimate. Fraudulent transactions do not follow a simple linear pattern but have a complex, nonlinear relationship. Which SVM kernel would be the most suitable?

- (a) Linear
- (b) Polynomial
- (c) Gaussian (RBF)
- (d) Sigmoid

Correct Answer: (c)
Explanation: The **RBF kernel** maps data into **higher-dimensional space**, making it **highly effective** for complex fraud patterns.

Question 11(MCQ)
A machine learning engineer is working on a **regression problem** and is concerned about the model's **complexity**. The engineer is considering using **regularization techniques** to control the model's complexity.
Which of the following is a **regularization technique**?
A) Linear Regression
B) Ridge Regression
C) Polynomial Regression
D) Logistic Regression
Correct Answer: B) Ridge Regression
Explanation: Ridge Regression adds **L2 regularization** to the loss function.

12. (MCQ) Topic: Classification | Level: Medium

Question: The Variance Inflation Factor (VIF) is a measure that quantifies multicollinearity in regression models. It indicates how much the variance of a regression coefficient is inflated due to correlations between independent variables.
A hospital is using multiple regression to predict patient recovery time based on various factors, including age, severity of illness, and treatment method. After training the model, the variance inflation factor (VIF) for some predictors is above 10. What should be done to improve the model?

- (a) Remove highly correlated independent variables.
- (b) Increase the sample size.
- (c) Add polynomial terms to the model.
- (d) Use k-means clustering instead.

Correct Answer: (a)
Explanation: A high VIF indicates multicollinearity. Removing highly correlated predictors can stabilize coefficient estimates.

13. (MCQ) Topic: Regularization | Level: Medium

Question: A pharmaceutical company develops a regression model to predict the effectiveness of a new drug based on dosage, patient age, and body mass index (BMI). They use Lasso regression and find that the coefficient of BMI has shrunk to zero. What does this imply?

- (a) BMI is highly correlated with other predictors and has been removed.
- (b) The model is underfitting due to excessive regularization.
- (c) BMI is the most important factor in predicting drug effectiveness.
- (d) Lasso regression does not work well with numerical variables.

Correct Answer: (a)
Explanation: Lasso regression applies L1 regularization, which can shrink some coefficients to zero, effectively removing them from the model.

Question 14

A news website has a "Terms of Service" page with **no outbound links**. The SEO team worries this might harm their PageRank distribution.

Question:
How will the PageRank algorithm handle this page's influence?
A) Discard its PageRank
B) Assign it the lowest rank
C) Distribute its rank equally to all pages
D) Let users decide via teleportation

Correct Answer: C) Distribute its rank equally to all pages
Explanation:
Pages with no outbound links (dangling nodes) have their PageRank redistributed evenly across all pages via the teleportation model.

15. (MCQ) Topic: Gradient Descent | Level: Medium

Question: A company is training a deep learning model using gradient descent. They observe that the loss function oscillates without decreasing significantly. What is the most likely reason for this behavior?

- (a) Learning rate is too high.
- (b) Learning rate is too low.
- (c) Model has too few training samples.
- (d) Too many features in the dataset.

Correct Answer: (a)
Explanation: A high learning rate causes updates to jump around the minimum instead of converging, leading to oscillations.

16. (MCQ) Topic: Gradient Descent | Level: Medium

Question: A company notices that training their gradient descent model is taking too long. What is the most likely cause?

- (a) Learning rate is too low.

- (b) The dataset is too small.
- (c) The model is overfitting.
- (d) The cost function is increasing.

Correct Answer: (a)
Explanation: A very small learning rate causes gradient descent to take small steps, making the learning process slow.

Question 17

A logistic regression model computes a **linear combination (z)** of input features and applies the **sigmoid function** to convert **z into a probability**. For a particular input, **z = 1**.

Question:

What is the probability using the sigmoid function? **(Round to 2 decimal places)**

Correct Answer: 0.73

To find the probability, substitute **z = 1** into the formula.

Explanation:
The **sigmoid function** is defined as:
1/ {1 + e^{-z}}

For z = 1,
1/ {1 + e^{-1}}

Thus, the **probability is 0.73** when **z = 1**.

18. (MCQ) Topic: Classification | Level: Medium

Question: A bank builds a binary classification model to detect fraud. The dataset contains **99% normal transactions** and **1% fraudulent transactions**. Which of the following issues is most likely to occur?

- (a) Overfitting.
- (b) Class imbalance.
- (c) Low recall.

- (d) High precision.

Correct Answer: (b)
Explanation: In imbalanced datasets, the model might learn to predict the majority class (normal transactions), leading to high accuracy but poor fraud detection.

19. (MCQ) Topic: Polynomial Regression | Level: Medium

Question: Ananya is working on a machine learning project to predict car prices based on features like engine power, mileage, and fuel type. She decides to use Polynomial Regression to capture potential non-linear relationships.

After training the model, she notices that it performs well on the training data but poorly on new test data. What is the most likely reason for this behavior?

(a) The model has too few polynomial terms, leading to underfitting.
(b) The model has too many polynomial terms, leading to overfitting.
(c) Polynomial regression does not work for numerical data.
(d) The model is not affected by the choice of polynomial degree.
Correct Answer: (b)
Explanation:
Polynomial regression can become too complex if the polynomial degree is too high, causing it to fit the noise in the training data rather than the underlying pattern. This results in overfitting, where the model has low training error but high test error. The solution is to regularize the model (e.g., Ridge Regression) or choose an appropriate polynomial degree using cross-validation.

Questions 20:
Scenario:
A data scientist is working on a **regression problem** and is considering using **regularization techniques** to improve the model's performance. The dataset has many features, and the model is showing signs of **overfitting**.

Which regularization technique adds the **absolute value of coefficients** to the loss function?
A) Ridge Regression
B) Lasso Regression
C) Elastic Net
D) Polynomial Regression

Correct Answer: B) Lasso Regression
Explanation: Lasso Regression adds the L1 regularization, which penalizes the sum of absolute values of coefficients, leading to feature selection

21. (MCQ) Topic: Classification | Level: Easy

Question: A company builds an email spam filter using classification. Which of the following would be the best choice for a target variable (Y)?

- (a) Email sender name
- (b) Number of words in the email
- (c) Whether the email is spam or not
- (d) The length of the email in characters

Correct Answer: (c)
Explanation: The target variable in classification should represent distinct categories, and here, the goal is to classify emails as **spam or not spam**.

22. (MSQ) Topic: Gradient Descent | Level: Medium

Question: Which of the following statements are **true** regarding gradient descent?

- (a) It minimizes the loss function.
- (b) It updates parameters w and b iteratively.
- (c) It can work without labeled data.
- (d) The learning rate α determines the step size.

Correct Answers: (a), (b), (d)
Explanation: Gradient descent is an **optimization algorithm** that minimizes a **loss function** by updating parameters step-by-step. It **requires labeled data** in supervised learning but not in unsupervised tasks.

23. (MCQ) Topic: Classification | Level: Medium

Question: An online magazine company wants to predict whether a **user will subscribe** to a paid plan after a **30-day free trial**. They collect data on:

- Reading frequency
- Time spent per visit
- Whether the user clicked on a promotional email

Which type of **machine learning approach** is best suited for this problem?

- (a) Clustering algorithm
- (b) Classification model
- (c) Regression model
- (d) Reinforcement learning

Correct Answer: (b)
Explanation: The outcome is **binary (Subscribe or Not Subscribe)**, making **classification** the best approach.

Question 24:

During a machine learning workshop, a participant uses PCA on a customer dataset. After calculating the eigenvalues, they decide to keep only the top two principal components. What is the most likely reason for this choice?

A. The top two eigenvalues always provide the most accurate results

B. PCA only works with two components

C. The top two eigenvectors represent directions that capture the most variance

D. It’s computationally easier to visualize 2D data

Answer: C

Explanation: Principal Component Analysis (PCA) transforms the dataset into a new coordinate system defined by the eigenvectors of the data’s covariance matrix. The associated eigenvalues represent the amount of variance captured by each direction. The directions (eigenvectors) with the highest eigenvalues correspond to the most informative components. Keeping the top two components helps simplify the data while retaining most of its structure.

25. (NAT) Topic: Gradient Descent | Level: Hard

Question:
A manufacturing company trains a predictive maintenance model using gradient descent. The cost function $J(w, b)$ is observed after every 10 iterations.

Iteration	Cost $J(w, b)$
0	10.0
10	5.2
20	2.8
30	1.5
40	1.0
50	0.95
60	0.94

The company decides to stop training when
J_prev - J_new < 0.001 . The algorithm will run for 60 iterations if stopping criteria is not met.
At which iteration should the training stop?

Correct Answer: 60
Explanation:
The difference between the 60th and 50th iterations is:

$$0.95 - 0.94 = 0.01$$

Since this value is greater than **0.001**, training **continues**. As per the question, the algorithm will only run for 60 iterations if the stopping criteria is not met.

26. (MSQ) Topic: SVM | Level: Medium

Question: Arjun is building an SVM model for image classification. He wants to ensure the model performs well and generalizes effectively. What does the effectiveness of an SVM mainly depend on?

- (a) Selection of Kernel Trick
- (b) Kernel Parameters

- (c) Learning Rate
- (d) All of the above

Correct Answer: (a) & (b) (Kernel Trick and Kernel Parameters)
Explanation:

- **Kernel Trick:** Helps SVM handle **non-linear data**.
- **Kernel Parameters:** Control the decision boundary's **flexibility**.
- **Learning Rate** is **not applicable** to SVM.

Question 27

Scenario:
A startup is building a search engine and uses the PageRank algorithm. The team sets a damping factor of 0.85 to prioritize hyperlink-based navigation over random jumps.

Question:
What percentage of user interactions will involve random jumps to any page in their 4-page prototype?
A) 5%
B) 15%
C) 25%
D) 85%

Correct Answer: B) 15%
Explanation:
Random jump probability = $(1-0.85)\times 100 = 15\%$

28. (MCQ) Topic: Polynomial Regression | Level: Medium

Question: Vikram is developing a predictive model to estimate stock prices based on past trends. He initially starts with a low-degree polynomial regression, but the predictions seem too simplistic. To improve accuracy, he keeps increasing the polynomial degree.

What is the likely issue Vikram will face if he increases the degree too much?

- (a) The model will always perform well on both training and test data.
- (b) The model may overfit the training data and perform poorly on new data.
- (c) The model will underfit the data.

- (d) The model will always be the best choice for any dataset.

Correct Answer: (b) The model may overfit the training data and perform poorly on new data.
Explanation: A **high-degree polynomial** can **overfit** the training data by capturing **noise**, leading to poor generalization on test data.

29. (MCQ) Topic: Polynomial Regression | Level: Hard

Question:
The learner is trying to predict the price of a house based on the **length** and **width** of the house. Let:

- **X_1** = Length
- **X_2** = Width

What could be a better hypothesis?

- (a) $h(X) = \Theta_0 + \Theta_1 X_1$
- (b) $h(X) = \Theta_0 + \Theta_2 X_2$
- (c) $h(X) = \Theta_0 + \Theta_2 X_1^2$
- (d) $h(X) = \Theta_0 + \Theta_1 Y$, where **Y (area of the house) =** $X_1 \times X_2$

Correct Answer: (d)
Explanation:
To predict the price of the house, the **size** is a better parameter. It can be determined by the **area of the house**, which is **length × width**. Instead of using the two features separately (X_1, X_2), a better single feature **Y = X_1 × X_2** can be used.

Question 30 (MCQ)

A logistic regression model is used to classify emails as spam or not spam. The model outputs a probability of 0.6 for a particular email.

Question: What is the decision rule in this case?

Options:

A) Predict 0

B) Predict 1

C) Predict 0.5

D) No decision

Correct Answer: B) Predict 1

Explanation:

- Logistic regression **outputs a probability** between **0 and 1**.
- The **default decision threshold** is **0.5**:
- If **probability ≥ 0.5**, predict **1** (spam).
- If **probability < 0.5**, predict **0** (not spam).