Dr. SNS RAJALAKSHMI COLLEGE OF ARTS AND SCIENCE (Autonomous)
Accredited by NAAC (Cycle-IV) with 'A+' Grade,
(Recognized by UGC & Approved by AICTE, New Delhi and Affiliated to Bharathiar University, Coimbatore)
486, Thudiyalur-Saravanampatti Road, Chinnavedampatti (Post), Coimbatore - 641 049.

Subject: DISCIPLINE CENTRIC ELECTIVE-3: MACHINE LEARNING IN CYBER SECURITY **Code:** 22UCB812

QUESTION AND ANSWER
UNIT: 1

1. Define the fundamental foundations of supervised learning and explain why labeled data is indispensable for training accurate predictive models. (Google / NET / 2025)

2. Explain the concept of inductive bias in decision trees and describe how it influences the selection of one hypothesis over another during training. (Amazon / CS / 2025)

3. Contrast the mathematical objectives of Linear Regression versus Logistic Regression in the context of continuous and categorical output variables. (Microsoft / NET / 2025)

4. Illustrate the trade-off between Regression and Classification by providing a specific cyber security example for each type of learning task. (Meta / Group Exams / 2024)

5. Summarize the process of Generalization in machine learning and explain why a model must perform well on unseen data to be considered effective. (Apple / NET / 2023)

6. Identify the core components of the training phase in supervised learning and list the steps required to minimize the error function. (Netflix / CS / 2025)

7. Describe the structural architecture of a basic Decision Tree and explain how information gain is used to determine the root node. (Oracle / CA / 2024)

8. Apply the concept of Linear Regression to a scenario where a security analyst needs to predict the growth rate of malware variants over time. (TCS / NET / 2025)

9. State the significance of supervised learning foundations in the development of automated intrusion detection systems for enterprise networks. (Infosys / CS / 2023)

10. Categorize the different types of learning algorithms and explain why supervised learning is often preferred for known threat detection. (Wipro / NET / 2024)

11. Describe the role of training data quality in preventing the model from learning noise instead of actual cyber threat patterns. (Google / CMA / 2025)

12. Show the relationship between inductive bias and the simplicity of the resulting model using the principle of Occam's Razor. (Amazon / Group Exams / 2024)

13. Contrast "Foundations of supervised learning" with unsupervised approaches regarding the requirement of a ground truth for performance evaluation. (Microsoft / CS / 2025)

14. Identify the primary differences between training error and generalization error in the context of building robust security classifiers. (Apple / NET / 2024)

15. Demonstrate the step-by-step construction of a Decision Tree using a small dataset of network traffic logs to classify normal versus malicious activity. (Google / CS / 2025)

16. Analyze how the "Inductive Bias" of a learning algorithm determines its ability to generalize from training examples to unseen instances. (Amazon / NET / 2025)

17. Evaluate the efficiency of Linear Regression versus Logistic Regression for a security task involving the prediction of the severity level of a vulnerability. (Microsoft / CMA / 2024)

18. Compare the foundations of supervised learning with the requirements of a typical cyber security pipeline for real-time threat detection. (TCS / CS / 2023)

19. Explain how "Generalization" ensures that a machine learning model does not simply memorize the training data but learns underlying security patterns. (Infosys / NET / 2025)

20. Analyze how a Decision Tree can be used to automate the generation of firewall rules based on historical traffic patterns. (Netflix / CS / 2024)

21. Assess the role of "Training" in minimizing the cost function of a Logistic Regression model used for binary malware classification. (Oracle / CS / 2025)

22. Illustrate the differences between Regression and Classification objectives by mapping each to a specific threat intelligence use case. (Apple / CA / 2025)

23. Critique the limitations of Decision Trees, such as their tendency to be unstable with small changes in data, and suggest ensemble solutions. (Meta / NET / 2025)

24. Formulate a set of features that would be essential for training a Linear Regression model to estimate the recovery time after a cyber attack. (Wipro / Group Exams / 2023)

25. Examine the impact of a strong inductive bias on the flexibility of a machine learning model when faced with highly varied cyber threats. (Google / CS / 2024)

26. Justify the need for supervised learning in cyber security environments where historical attack data is well-documented and labeled. (Microsoft / CMA / 2025)

27. Categorize the common challenges in "Generalization and Training" when the distribution of attack data changes over time (concept drift). (Apple / NET / 2024)

28. Implement a logical diagram showing the workflow of supervised learning from data collection and labeling to model deployment and monitoring. (Amazon / CS / 2025)

29. Design a comprehensive machine learning pipeline for a new security startup, covering the foundations of supervised learning and inductive bias. (Google / NET / 2025)

30. Analyze the mathematical relationships between feature selection, inductive bias, and the resulting generalization performance of a security classifier. (Amazon / CS / 2025)

31. Evaluate the comparative effectiveness of Decision Trees, Linear Regression, and Logistic Regression for modeling different aspects of a DDoS attack. (Microsoft / CMA / 2025)

32. Case Study (Anomaly Detection): Design a multi-stage supervised learning system to identify and classify sophisticated malware based on their behavioral signatures. (Apple / CS / 2024)

33. Formulate a set of best practices for ensuring that a machine learning model for threat detection maintains high performance during the training phase. (TCS / NET / 2025)

34. Examine the future trends in supervised learning for cyber security, focusing on how "inductive bias" might be automated via meta-learning. (Infosys / Group Exams / 2023)

35. Propose a technical roadmap for transitioning a traditional rule-based firewall to a fully automated system driven by Logistic Regression and Decision Trees. (Wipro / NET / 2025)

36. Critique the reliance on labeled datasets for supervised learning in security and propose semi-supervised methods to leverage the abundance of unlabeled logs. (Oracle / CS / 2024)

Dr. SNS RAJALAKSHMI COLLEGE OF ARTS AND SCIENCE (Autonomous)
Accredited by NAAC (Cycle-IV) with 'A+' Grade,
(Recognized by UGC & Approved by AICTE, New Delhi and Affiliated to Bharathiar University, Coimbatore)
486, Thudiyalur-Saravanampatti Road, Chinnavedampatti (Post), Coimbatore - 641 049.

Subject: DISCIPLINE CENTRIC ELECTIVE-3: MACHINE LEARNING IN CYBER SECURITY **Code:** 22UCB812

QUESTION AND ANSWER
UNIT: 2

1. Define the problem of Overfitting in machine learning and explain how it leads to poor performance on new, independent datasets. (Amazon / NET / 2025)

2. Explain the Bias vs Variance dilemma and describe how increasing model complexity typically affects these two sources of error. (Google / CS / 2024)

3. List the essential Performance metrics used for evaluating classifiers and define the importance of the F1-score in imbalanced security datasets. (Microsoft / NET / 2025)

4. Describe the operational mechanics of a Random Forest and explain how the aggregation of multiple trees improves prediction accuracy. (Apple / Group Exams / 2023)

5. Illustrate the functionality of a Perceptron as a binary classifier and explain the role of the activation function in producing a discrete output. (Meta / CS / 2024)

6. Analyze the challenges of "Beyond binary classification" when a security system needs to categorize threats into multiple distinct families. (Netflix / NET / 2025)

7. State the purpose of a Validation set and explain how it differs from a Test set in the overall model development lifecycle. (Oracle / CA / 2025)

8. Apply Anomaly Detection techniques to a case study involving the identification of unusual traffic patterns in a corporate local area network. (TCS / CS / 2024)

9. Summarize the benefits of using ensemble methods like Random Forest over a single Decision Tree for complex cyber security datasets. (Infosys / NET / 2023)

10. Contrast the concept of "Bias" with "Variance" by showing how each contributes to total prediction error in a machine learning model. (Wipro / CMA / 2025)

11. Identify the key features of the Perceptron learning rule and explain why it is considered the building block of neural networks. (Google / Group Exams / 2024)

12. Describe how "Performance metrics" such as Precision and Recall help in assessing the risk of false positives in threat detection. (Amazon / NET / 2025)

13. Show the flow of the "Validation and Testing" phase and explain how it helps in selecting the best hyperparameters for a model. (Microsoft / CS / 2024)

14. Distinguish between a model that is "Underfitting" and one that is "Overfitting" based on their training and testing accuracy scores. (Apple / NET / 2025)

15. Demonstrate how to perform "Validation and Testing" on a Random Forest model to ensure it is not overfitting to a specific malware dataset. (Google / CS / 2025)

16. Analyze the trade-offs between Bias and Variance in a Perceptron model and describe how to find the optimal balance for high accuracy. (Amazon / NET / 2025)

17. Evaluate the effectiveness of different Performance metrics (Accuracy vs AUC) for an anomaly detection system where threats are very rare. (Microsoft / CMA / 2024)

18. Compare a single Decision Tree with a Random Forest in terms of their robustness against noise and their tendency to overfit. (TCS / CS / 2023)

19. Explain the logic of Anomaly Detection and discuss why it is critical for identifying "zero-day" attacks that have no existing signatures. (Infosys / NET / 2025)

20. Analyze how a Perceptron-based system can be trained to distinguish between authorized and unauthorized access attempts. (Amazon / CS / 2024)

21. Assess the utility of "Beyond binary classification" for a security system that needs to label files as Benign, Spyware, Ransomware, or Adware. (Oracle / CS / 2025)

22. Illustrate the concept of Overfitting using a learning curve that plots training and validation error against the number of training iterations. (Apple / CA / 2025)

23. Examine the importance of choosing appropriate performance metrics for a cyber security model where a False Negative is much costlier than a False Positive. (Meta / NET / 2025)

24. Justify the choice of using Random Forest for feature importance ranking in a dataset containing hundreds of network traffic attributes. (Netflix / Group Exams / 2023)

25. Calculate the impact of high variance on the reliability of a security classifier when it is deployed across different network environments. (Google / CS / 2024)

26. Categorize the various "Validation and Testing" strategies, such as K-fold cross-validation, and explain their role in reducing selection bias. (Apple / CMA / 2025)

27. Propose a strategy for mitigating overfitting in a deep Decision Tree without reducing the depth of the tree itself (e.g., using pruning). (Microsoft / NET / 2024)

28. Demonstrate the process of setting up an Anomaly Detection pipeline that uses statistical thresholding to trigger alerts for high-volume traffic. (Amazon / CS / 2025)

29. Design a robust "Validation and Testing" framework that uses stratified cross-validation and multiple performance metrics to evaluate a new ransomware detector. (Google / NET / 2025)

30. Analyze the impact of "Overfitting" and the "Bias vs Variance" trade-off on the reliability of machine learning models in critical security infrastructure. (Amazon / CS / 2025)

31. Evaluate the implementation of an ensemble model (Random Forest) versus a simple Perceptron for identifying zero-day exploits in a large enterprise. (Microsoft / CMA / 2025)

32. Design an anomaly detection system that uses a combination of Perceptrons and Decision Trees to flag suspicious behavior. (Meta / CS / 2024)

33. Formulate a comprehensive performance audit for a security model that includes Precision-Recall curves, ROC-AUC, and Confusion Matrices. (TCS / NET / 2025)

34. Examine the role of "Beyond binary classification" in achieving a more granular and actionable threat intelligence report for a security operations center. (Infosys / Group Exams / 2023)

35. Propose a "Validation and Testing" protocol for models deployed in highly dynamic environments where the underlying data distributions change hourly. (Wipro / NET / 2025)

36. Critique the use of Accuracy as a primary metric in cyber security and suggest more robust alternatives that account for the high cost of false negatives. (Oracle / CS / 2024)

Dr. SNS RAJALAKSHMI COLLEGE OF ARTS AND SCIENCE (Autonomous)
Accredited by NAAC (Cycle-IV) with 'A+' Grade,
(Recognized by UGC & Approved by AICTE, New Delhi and Affiliated to Bharathiar University, Coimbatore)
486, Thudiyalur-Saravanampatti Road, Chinnavedampatti (Post), Coimbatore - 641 049.

Subject: DISCIPLINE CENTRIC ELECTIVE-3: MACHINE LEARNING IN CYBER SECURITY **Code:** 22UCB812

QUESTION AND ANSWER
UNIT: 3

1. Define the Naive Bayes algorithm and explain the "independence assumption" that simplifies the calculation of posterior probabilities. (Meta / NET / 2025)

2. List the components of a Bayesian Belief Network and describe how it represents conditional dependencies among a set of random variables. (Google / CS / 2024)

3. Explain the logic of the K-Nearest Neighbour (KNN) algorithm and describe how the choice of 'K' affects the smoothness of decision boundaries. (Amazon / NET / 2025)

4. Describe the geometric intuition behind Support Vector Machines (SVM) and explain the concept of finding the optimal separating hyperplane. (Microsoft / Group Exams / 2024)

5. Apply the Naive Bayes classifier to a simple cyber security scenario, such as determining if an incoming email is likely to be phishing. (TCS / CA / 2025)

6. Contrast KNN with SVM in terms of computational complexity during the testing phase for large-scale security log analysis. (Infosys / CS / 2023)

7. Identify the role of "Kernels" in Support Vector Machines and explain how they allow for the classification of non-linearly separable data. (Wipro / NET / 2025)

8. Summarize the advantages of Bayesian Belief Networks in modeling complex security threats where multiple interconnected factors are involved. (Oracle / Group Exams / 2024)

9. Illustrate how K-Nearest Neighbour can be used for outlier detection in a dataset containing user login timestamps and locations. (Apple / CS / 2025)

10. Analyze the sensitivity of the Naive Bayes classifier to irrelevant features and explain how this impacts its use in feature-heavy datasets. (Netflix / NET / 2024)

11. State the significance of the "Maximum Margin" in Support Vector Machines for improving the robustness of the resulting classification. (Google / NET / 2023)

12. Describe the process of training a Bayesian Belief Network from data and explain how prior knowledge can be incorporated into the model. (Amazon / CS / 2025)

13. Show the decision-making process of an SVM when presented with data points that lie very close to the decision boundary. (Microsoft / CMA / 2025)

14. Distinguish between parametric and non-parametric models using Naive Bayes and KNN as respective examples for each category. (Meta / NET / 2024)

15. Demonstrate the application of a Naive Bayes classifier for real-time website categorization to prevent users from accessing malicious URLs. (Google / CS / 2025)

16. Analyze the structural dependencies in a Bayesian Belief Network designed to model the likelihood of a multi-stage data breach. (Amazon / NET / 2025)

17. Evaluate the performance of K-Nearest Neighbour for detecting anomalous user behavior in a system with thousands of active sessions. (Microsoft / CMA / 2024)

18. Compare the computational efficiency of Naive Bayes and Support Vector Machines for high-frequency network packet classification. (TCS / CS / 2023)

19. Explain the mechanism of SVM and how the choice of the kernel function impacts the model's ability to create complex decision boundaries. (Infosys / NET / 2025)

20. Analyze how a Bayesian Belief Network can integrate multiple indicators like sender reputation and content analysis. (Wipro / Group Exams / 2024)

21. Assess the utility of the "Maximum Margin" concept in SVM for ensuring that a classifier remains robust against adversarial noise. (Oracle / CS / 2025)

22. Illustrate the process of selecting the optimal value of 'K' in a KNN algorithm using a validation set and a range of candidate values. (Apple / CA / 2025)

23. Examine how Naive Bayes handles missing data during the prediction phase and explain why this property is useful for sparse security logs. (Meta / NET / 2025)

24. Justify the use of Bayesian Belief Networks for "Root Cause Analysis" after a security incident has been detected in a distributed system. (Netflix / Group Exams / 2023)

25. Calculate the impact of the "Naive" assumption in Naive Bayes on the final probability estimates when features are highly correlated. (Google / CS / 2024)

26. Categorize various types of SVM kernels (Linear, Polynomial, RBF) and identify the specific data characteristics where each is most appropriate. (Apple / CMA / 2025)

27. Propose a hybrid security model that combines the speed of Naive Bayes for initial filtering and the accuracy of SVM for final verification. (Microsoft / NET / 2024)

28. Demonstrate how to resolve the "Cold Start" problem in a KNN-based threat detection system when no historical neighbors are available. (Amazon / CS / 2025)

29. Design a multi-layered classification strategy that uses Naive Bayes for fast triage and SVM for deep analysis of suspicious file attachments. (Google / NET / 2025)

30. Analyze the mathematical correlation between the kernel trick in SVM and the ability to find linear separators in a higher-dimensional feature space. (Amazon / CS / 2025)

31. Evaluate the trade-offs between KNN, Naive Bayes, and Bayesian Belief Networks for modeling the probability of an internal data theft incident. (Microsoft / CMA / 2025)

32. Design a Bayesian Belief Network that incorporates factors like domain age, SSL status, and content similarity. (Netflix / CS / 2024)

33. Formulate a strategy for scaling a KNN-based threat detection system to handle millions of new logs daily without compromising on query latency. (TCS / NET / 2025)

34. Examine the security implications of "Adversarial Machine Learning" where attackers craft inputs specifically designed to bypass an SVM classifier. (Infosys / Group Exams / 2023)

35. Propose a data governance framework for Naive Bayes models that includes automated feature engineering and regular recalibration of prior probabilities. (Wipro / NET / 2025)

36. Critique the use of Naive Bayes for tasks with highly dependent features and suggest architectural changes to incorporate conditional dependencies (e.g., BBN). (Oracle / CS / 2024)

Dr. SNS RAJALAKSHMI COLLEGE OF ARTS AND SCIENCE (Autonomous)
Accredited by NAAC (Cycle-IV) with 'A+' Grade,
(Recognized by UGC & Approved by AICTE, New Delhi and Affiliated to Bharathiar University, Coimbatore)
486, Thudiyalur-Saravanampatti Road, Chinnavedampatti (Post), Coimbatore - 641 049.

Subject: DISCIPLINE CENTRIC ELECTIVE-3: MACHINE LEARNING IN CYBER SECURITY **Code:** 22UCB812

QUESTION AND ANSWER
UNIT: 4

1. Define a Markov Model and explain the fundamental Markov property where the future state depends only on the current state. (Amazon / NET / 2025)

2. List the elements of a Hidden Markov Model (HMM) and explain why it is useful for modeling processes with unobservable states. (Google / CS / 2024)

3. Explain the difference between Maximum Likelihood Estimation (MLE) and Bayesian Estimate for parameter estimation in statistical models. (Microsoft / NET / 2025)

4. Describe the two main steps of the Expectation Maximization (EM) algorithm and explain how it handles datasets with missing variables. (Apple / Group Exams / 2023)

5. Apply Neural Networks to a scenario involving the recognition of malicious patterns in high-dimensional network traffic data. (Meta / CS / 2024)

6. Contrast Markov Models with Hidden Markov Models in terms of state visibility and the types of problems they are designed to solve. (Netflix / NET / 2025)

7. Identify the role of weights and biases in a single-layer Neural Network and explain how they are updated during backpropagation. (Oracle / CA / 2025)

8. Summarize the advantages of using Bayesian Estimation when dealing with small datasets where prior information is available. (TCS / CS / 2024)

9. Illustrate the architecture of a multi-layer Neural Network, including input, hidden, and output layers, and describe their interactions. (Infosys / NET / 2023)

10. Analyze the computational complexity of the Expectation Maximization algorithm when applied to large-scale security log clustering. (Wipro / CMA / 2025)

11. State the purpose of the "Transition Matrix" in a Markov Model and explain how it defines the probability of moving between states. (Google / Group Exams / 2024)

12. Describe how Hidden Markov Models can be used to detect sophisticated multi-stage cyber attacks that occur over long periods. (Amazon / NET / 2025)

13. Show the process of parameter estimation using MLE for a dataset following a normal distribution of network packet sizes. (Microsoft / CS / 2024)

14. Examine the impact of the learning rate on the convergence speed and stability of a Neural Network during the training phase. (Apple / NET / 2025)

15. Demonstrate the setup of a Hidden Markov Model to track the progression of a user's behavior across different security states (Normal, Suspect, Attack). (Google / CS / 2025)

16. Analyze the internal structure of a multi-layer Neural Network and how it facilitates the learning of complex, non-linear relationships in security data. (Amazon / NET / 2025)

17. Evaluate the performance differences between MLE and Bayesian Estimate when estimating the parameters of a model with very limited attack samples. (Microsoft / CMA / 2024)

18. Compare the operational overhead of a Markov Model versus a Neural Network for a simple network traffic classification task. (TCS / CS / 2023)

19. Explain the importance of the Expectation Maximization algorithm in training models where some of the state information is hidden or unrecorded. (Infosys / NET / 2025)

20. Analyze how a multi-layer Neural Network can be trained to detect coordinated botnet activity across multiple hosts. (Meta / Group Exams / 2024)

21. Assess the role of "Hidden States" in an HMM for modeling the internal logic of a sophisticated, stealthy malware variant. (Oracle / CS / 2025)

22. Illustrate the steps involved in the "E-step" and "M-step" of the Expectation Maximization algorithm using a simple clustering example. (Apple / CA / 2025)

23. Examine the differences between a First-order Markov Model and a Higher-order model in terms of memory and prediction accuracy. (Meta / NET / 2025)

24. Justify the choice of using Neural Networks for automated security log parsing and classification over traditional rule-based systems. (Netflix / Group Exams / 2023)

25. Calculate the impact of varying the prior probability in a Bayesian Estimate on the final posterior distribution of attack likelihood. (Google / CS / 2024)

26. Categorize the common activation functions used in Neural Networks (Sigmoid, ReLU, Tanh) and explain their impact on vanishing gradients. (Apple / CMA / 2025)

27. Propose a Markov-based strategy for predicting the next likely step in a multi-stage attack based on observed security events. (Microsoft / NET / 2024)

28. Demonstrate how to set up a Neural Network with dropout layers to prevent overfitting on a dataset containing noisy network traffic. (Amazon / CS / 2025)

29. Design a sophisticated intrusion detection system that combines Hidden Markov Models for temporal analysis and Neural Networks for spatial feature extraction. (Google / NET / 2025)

30. Analyze the cost-performance ratio of using Bayesian Estimation versus MLE for training security models in data-scarce environments like zero-day detection. (Amazon / CS / 2025)

31. Evaluate the architectural limitations of Neural Networks, specifically regarding "Explainability" and "Transparency" in high-stakes security decision making. (Microsoft / CMA / 2025)

32. Design an HMM-based system that can correctly identify the phase of an Advanced Persistent Threat (APT). (Amazon / CS / 2024)

33. Formulate a comprehensive training strategy for a deep Neural Network that includes proper weight initialization, batch normalization, and early stopping. (TCS / NET / 2025)

34. Examine the internal mechanics of the Expectation Maximization algorithm and how it can be used to uncover latent variables in encrypted traffic. (Infosys / Group Exams / 2023)

35. Propose a hybrid model that uses Markov chains to generate synthetic attack data for training more robust Neural Network security classifiers. (Wipro / NET / 2025)

36. Critique the "Black Box" nature of Neural Networks and suggest methods like Layer-wise Relevance Propagation to provide interpretable security alerts. (Oracle / CS / 2024)

Dr. SNS RAJALAKSHMI COLLEGE OF ARTS AND SCIENCE (Autonomous)
Accredited by NAAC (Cycle-IV) with 'A+' Grade,
(Recognized by UGC & Approved by AICTE, New Delhi and Affiliated to Bharathiar University, Coimbatore)
486, Thudiyalur-Saravanampatti Road, Chinnavedampatti (Post), Coimbatore - 641 049.

Subject: DISCIPLINE CENTRIC ELECTIVE-3: MACHINE LEARNING IN CYBER SECURITY **Code:** 22UCB812

QUESTION AND ANSWER
UNIT: 5

1. Define the "Curse of Dimensionality" and explain why high-dimensional data poses significant challenges for traditional machine learning algorithms. (Meta / NET / 2025)

2. List the primary goals of Dimensionality Reduction Techniques and explain how they help in visualizing and simplifying complex datasets. (Google / CS / 2024)

3. Explain the mathematical foundation of Principal Component Analysis (PCA) and describe how it identifies the directions of maximum variance. (Amazon / NET / 2025)

4. Describe the logic of K-means clustering and explain how the algorithm iteratively assigns data points to the nearest cluster center. (Microsoft / Group Exams / 2024)

5. Apply Linear Discriminant Analysis (LDA) to a dimensionality reduction problem where class labels are available to maximize class separability. (TCS / CA / 2025)

6. Contrast K-means clustering with Hierarchical clustering in terms of their approach to building cluster structures (flat vs nested). (Infosys / CS / 2023)

7. Identify the core concepts of Association Rule Mining and define the metrics of Support, Confidence, and Lift used in the process. (Wipro / NET / 2025)

8. Summarize the functionality of Spectral clustering and explain why it is more effective than K-means for non-convex cluster shapes. (Oracle / Group Exams / 2024)

9. Illustrate a case study of using machine learning for Spam Filtering and describe the types of features typically extracted from emails. (Apple / CS / 2025)

10. Analyze the benefits of Subspace Clustering for datasets where different clusters may exist in different subsets of features. (Netflix / NET / 2024)

11. State the role of Principal Component Analysis in reducing noise and redundant information in cyber security telemetry data. (Google / NET / 2023)

12. Describe how Hierarchical clustering produces a dendrogram and explain how this visualization helps in determining the number of clusters. (Amazon / CS / 2025)

13. Show the application of Machine Learning for End Point Protection in a case study involving real-time threat detection on user devices. (Microsoft / CMA / 2025)

14. Distinguish between PCA and LDA by highlighting that one is an unsupervised technique while the other is a supervised technique. (Meta / NET / 2024)

15. Demonstrate the application of PCA for reducing the dimensionality of a dataset containing thousands of binary malware features. (Google / CS / 2025)

16. Analyze the architecture of K-means clustering and how it can be used to group similar security alerts together for analyst review. (Amazon / NET / 2025)

17. Evaluate the advantages of using Linear Discriminant Analysis over PCA when class labels are available for enhancing threat classification. (Microsoft / CMA / 2024)

18. Compare Hierarchical clustering with Spectral clustering in terms of their ability to handle non-spherical clusters in security data. (TCS / CS / 2023)

19. Explain the workflow of Association Rule Mining and how it can reveal hidden correlations between different types of security events. (Infosys / NET / 2025)

20. Case Study (Spam Filtering): Analyze how unsupervised clustering can identify new families of spam emails without prior training on those specific types. (Netflix / Group Exams / 2024)

21. Assess the role of "Curse of Dimensionality" in making the Euclidean distance metric unreliable for high-dimensional security log analysis. (Oracle / CS / 2025)

22. Illustrate the setup of an Association Rule Mining task to find sets of security rules that frequently trigger together on a network. (Apple / CA / 2025)

23. Examine the importance of choosing between flat and hierarchical clustering based on the need for multi-level threat categorization. (Meta / NET / 2025)

24. Justify the use of Subspace Clustering for a dataset where relevant attack patterns are only visible in a specific subset of features. (Netflix / Group Exams / 2023)

25. Calculate the reduction in data volume when applying PCA to retain 95% of the variance in a 1000-feature cyber security dataset. (Google / CS / 2024)

26. Categorize various "Dimensionality Reduction Techniques" and their role in improving the performance of downstream classification models. (Apple / CMA / 2025)

27. Propose a clustering strategy for "Machine Learning for End Point Protection" that identifies previously unknown malware clusters on user machines. (Microsoft / NET / 2024)

28. Demonstrate how to configure a K-means algorithm to find the optimal number of clusters using the "Elbow Method" on security log data. (Amazon / CS / 2025)

29. Design an end-to-end unsupervised security system that uses PCA for noise reduction and Spectral clustering for identifying new botnet families. (Google / NET / 2025)

30. Analyze the mathematical differences between K-means, Hierarchical, and Subspace clustering and their respective strengths in high-dimensional security data. (Amazon / CS / 2025)

31. Evaluate the scalability differences between PCA and LDA for a real-time stream of millions of network packets per second. (Microsoft / CMA / 2025)

32. Design an unsupervised system that uses Association Rule Mining to identify common sequences of malicious system calls. (Meta / CS / 2024)

33. Formulate a dimensionality reduction strategy that combines PCA and LDA to maximize the interpretability and class separability of malware clusters. (TCS / NET / 2025)

34. Examine the importance of the "Curse of Dimensionality" in modern cyber security and suggest strategies beyond PCA to mitigate its effects. (Infosys / Group Exams / 2023)

35. Propose a "Zero Trust" security architecture that uses Spectral clustering to continuously verify user behavior against their historical baseline. (Wipro / NET / 2025)

36. Critique the use of K-means for security clustering where clusters may be of very different sizes and suggest more adaptive alternatives like DBSCAN. (Oracle / CS / 2024)