Concept of Machine Learning (CSA-202)

Question Bank

Unit 1: Core Concepts of Machine Learning

1. What is machine learning, and how does it differ from traditional programming?
2. What are the main types of machine learning algorithms?
3. How can machine learning be used to solve real-world problems?
4. Explain the concept of a “machine learning mindset.”
5. What are the key steps involved in the machine learning pipeline?
6. How would you define a machine learning problem? Provide an example.
7. What makes a problem suitable for machine learning?
8. Describe an example of a hard machine learning problem.
9. What is problem framing in machine learning, and why is it important?
10. What are some common machine learning problems encountered in different industries?
11. How is machine learning applied in image recognition?
12. Describe the role of machine learning in speech recognition.
13. Explain how machine learning can be used in medical diagnosis.
14. What is statistical arbitrage, and how does machine learning contribute to it?
15. How does machine learning help in learning associations between different data sets?
16. Describe the use of machine learning in the retail industry.
17. What is the significance of machine learning in fraud detection?
18. What is the difference between classification and regression in machine learning?
19. Explain how clustering is used in machine learning.
20. What is dimensionality reduction, and why is it important?
21. Define ranking as a machine learning task. Provide an example.
22. What is the difference between supervised and unsupervised learning?
23. What are the key challenges faced when using reinforcement learning?
24. How is the concept of manifold learning related to dimensionality reduction?
25. What is the importance of feature selection in the machine learning pipeline?
26. What are features in a machine learning model, and how are they selected?
27. Define what is meant by labels in supervised learning.
28. What are hyperparameters, and how do they differ from model parameters?
29. Why is it necessary to have a validation dataset in machine learning?
30. Explain the role of the test set in evaluating machine learning models.
31. What is a loss function, and why is it critical in training models?
32. How do hypothesis tests apply to machine learning?
33. What is supervised learning, and how is it applied in real-world problems?
34. How does unsupervised learning differ from supervised learning?
35. What is reinforcement learning, and what are its key components?
36. Explain semi-supervised learning and its use cases.
37. Define transductive inference and provide an example.
38. What is the role of online learning in machine learning?
39. How does active learning improve model performance with less data?
40. What is generalization in machine learning, and why is it crucial?

41. What is feature engineering, and why is it critical to machine learning success?
42. Describe different types of data that are typically used in machine learning.
43. How does big data influence the performance of machine learning models?
44. Why is data quality important in machine learning?
45. What are the major tasks involved in data preprocessing?
46. What methods are used to handle missing values in a dataset?
47. How do you address noisy data in machine learning?
48. Explain the data cleaning process and its importance.
49. What is the entity identification problem in data integration?
50. How does redundancy affect machine learning models, and how can it be resolved?
51. What is tuple duplication, and how does it impact data quality?
52. Explain how data value conflicts are detected and resolved.
53. What are the major data reduction strategies used in machine learning?
54. How is attribute subset selection performed in data reduction?
55. What are histograms, and how are they used in machine learning?
56. How does clustering contribute to data reduction?
57. What role does sampling play in data reduction?
58. Explain how data cube aggregation is used for data reduction.
59. What are the main strategies for data transformation in machine learning?
60. Describe the process of data discretization and its importance in machine learning.

Unit 2: Supervised Learning Algorithms - Part One

1. What is the bias-variance trade-off in supervised learning?
2. How does the complexity of a function affect the bias-variance trade-off?
3. Explain the relationship between the amount of training data and model performance.
4. What is meant by the dimensionality of the input space, and why is it significant in supervised learning?
5. How does noise in the output values affect model learning?
6. Why is it important to consider heterogeneity in the data when choosing an algorithm?
7. How can redundancy in the data affect the performance of a supervised learning model?
8. Explain the impact of non-linearities and interactions between features on the learning process.
9. What are the key factors that should be considered when selecting a supervised learning algorithm?
10. How does the number of training examples influence the generalization of a model?
11. What is the mathematical representation of a simple linear regression model?
12. How is the relationship between the dependent and independent variables represented in linear regression?
13. What assumptions must be made when using a linear regression model?
14. How does linear regression handle multiple input variables (multiple linear regression)?
15. Explain the significance of the slope and intercept in a linear regression model.
16. What role does the error term play in the linear regression model?
17. What is ordinary least squares (OLS) in linear regression?
18. How does OLS help in minimizing the cost function in linear regression?

19. What is gradient descent, and how is it used to learn the parameters in linear regression?
20. Explain the difference between batch gradient descent and stochastic gradient descent.
21. Describe the process of simple linear regression.
22. How do you interpret the results of a fitted linear regression model?
23. What is the purpose of regularization in supervised learning?
24. Explain how Lasso regression works and its application in model selection.
25. What is Ridge regression, and how does it differ from Lasso regression?
26. How does regularization help address the issue of overfitting?
27. Explain the bias-variance trade-off in the context of regularization.
28. What is the difference between L1 and L2 regularization techniques?
29. How is the cost function defined in linear regression, and why is it important?
30. What is feature scaling, and why is it necessary in gradient-based learning algorithms?
31. Explain the difference between normalization and mean normalization in machine learning.
32. How is the learning rate used in gradient descent, and how does it impact model convergence?
33. What is an automatic convergence test in gradient descent, and how does it work?
34. How do you evaluate the performance of a linear regression model?
35. What is logistic regression, and how does it differ from linear regression?
36. Explain the logistic function and its mathematical representation.
37. How are odds, odds ratios, and logit interpreted in logistic regression?
38. What is the inverse of the logistic function, and how is it used in predictions?
39. How does logistic regression handle multiple explanatory variables?
40. What is the "latent variable" interpretation of the logistic model?
41. Explain the relationship between the odds ratio and the logit in logistic regression.
42. What is the "rule of ten" in logistic regression, and why is it important?
43. How does the iteratively reweighted least squares (IRLS) algorithm work in logistic regression?
44. What methods are used to evaluate the goodness of fit in logistic regression models?
45. What are the limitations of logistic regression when applied to real-world problems?
46. What is linear discriminant analysis, and how does it differ from logistic regression?
47. Explain the concept of discriminant functions in LDA.
48. How does LDA work for two-class classification problems?
49. What assumptions are made when using LDA for classification?
50. How are eigenvalues used in linear discriminant analysis?
51. What is a discrimination rule in LDA, and how is it applied to classification?
52. What is the effect size in LDA, and how is it calculated?
53. How can linear discriminant analysis be applied in bankruptcy prediction?
54. Explain how LDA is used in face recognition systems.
55. What are the main applications of logistic regression in biomedical studies?
56. How can logistic regression be used in marketing analysis?
57. What is the practical significance of logistic regression in the field of machine learning?
58. How does linear discriminant analysis compare with logistic regression in terms of assumptions?
59. In what situations would you prefer LDA over logistic regression, and why?
60. How do LDA and logistic regression differ in terms of their ability to handle non-linearly separable data?

Unit 3: Supervised Learning Algorithms - Part Two

1. What is a Support Vector Machine (SVM), and how is it used in classification?
2. Explain the concept of a hyperplane in SVM.
3. What is the difference between a hard-margin and a soft-margin SVM?
4. What are support vectors in the context of SVM?
5. How does a hard-margin SVM work, and what are its limitations?
6. What is a soft-margin SVM, and how does it improve upon hard-margin SVM?
7. How does SVM handle non-linear classification problems?
8. What is the kernel trick in SVM, and why is it important?
9. What are common kernel functions used in SVMs (e.g., polynomial, RBF)?
10. Explain the primal form of the SVM optimization problem.
11. What is the dual form of the SVM optimization problem, and why is it used?
12. Describe how the kernel trick is used to transform the SVM into a non-linear classifier.
13. How does sub-gradient descent work in solving SVM optimization problems?
14. Explain the concept of coordinate descent in SVM optimization.
15. What is empirical risk minimization in the context of SVM?
16. How does SVM perform risk minimization?
17. What is regularization, and how does it help with the stability of an SVM model?
18. Explain the hinge loss function in SVM and its importance.
19. What is the target function in SVM, and how is it optimized?
20. What are the main parameters that need to be selected when training an SVM?
21. What challenges arise in choosing the appropriate kernel for an SVM?
22. How do you tune the regularization parameter C in a soft-margin SVM?
23. What are the common issues encountered while using SVM in practice?
24. What is an artificial neural network, and what are its basic components?
25. How does a feed-forward neural network function?
26. What are the weight-space symmetries in a neural network?
27. Explain the concept of activation functions in a neural network.
28. What are the common activation functions used in ANNs?
29. How is the output of a neural network computed given an input?
30. What is the difference between a single-layer and multi-layer perceptron?
31. What is the objective of training a neural network?
32. How is parameter optimization performed during neural network training?
33. Explain the local quadratic approximation used in neural networks.
34. How is gradient information used during neural network training?
35. What is gradient descent, and how is it applied in neural network training?
36. What are the differences between batch gradient descent and stochastic gradient descent?
37. How do you determine when to stop training a neural network?
38. What is backpropagation in neural networks?
39. How are error-function derivatives evaluated in backpropagation?
40. Provide a simple example of error backpropagation in a small network.
41. Why is backpropagation considered an efficient algorithm for training ANNs?
42. How does the learning rate affect the backpropagation algorithm?
43. What is a decision tree, and how is it used for classification?

44. How is a decision tree represented in machine learning?
45. Explain the ID3 learning algorithm used to build decision trees.
46. What is entropy in the context of decision tree learning?
47. How is information gain calculated in decision trees?
48. What does it mean for a decision tree to overfit the data?
49. What are common validation methods used to evaluate decision trees?
50. How can overfitting be avoided in decision tree learning?
51. What is the minimum-description length method, and how does it apply to decision trees?
52. How does noise in the data affect decision tree learning?
53. What is the random forest algorithm, and how is it related to decision trees?
54. How does bagging improve the performance of decision trees in a random forest?
55. What is the difference between bagging and boosting in ensemble methods?
56. How do random forests use the concept of random feature selection?
57. What are "extra trees," and how do they differ from regular trees in a random forest?
58. How is the variable importance measured in a random forest model?
59. What are the main advantages of using random forests over individual decision trees?
60. How do you interpret the results of a random forest model, particularly in terms of feature importance?

Unit 4: Unsupervised Learning

1. What is unsupervised learning, and how does it differ from supervised learning?
2. What are the key challenges faced in unsupervised learning?
3. What types of problems can be solved using unsupervised learning techniques?
4. What are the main applications of unsupervised learning in the real world?
5. Explain the concept of clustering in unsupervised learning.
6. What is the main goal of clustering in data analysis?
7. What are clustering methods based on Euclidean distance, and how do they work?
8. Explain the concept of clustering methods based on probabilities.
9. What are the key differences between hierarchical clustering and partition-based clustering methods?
10. What is hierarchical clustering, and how does it differ from other clustering techniques?
11. How does the choice of distance metric (e.g., Euclidean distance) impact clustering results?
12. What is the importance of the linkage criterion in hierarchical clustering?
13. What are the advantages and disadvantages of hierarchical clustering?
14. How do clustering methods based on probabilities differ from distance-based methods?
15. In what situations would a probabilistic clustering method be preferable to a distance-based method?
16. What is a dendrogram, and how is it used in hierarchical clustering?
17. What is the k-means clustering algorithm, and how does it work?
18. How is the number of clusters (k) selected in the k-means algorithm?
19. What are the steps involved in the naive (standard) k-means algorithm?

20. Explain the concept of cluster centroids in the k-means algorithm.
21. What is the purpose of the "assigning step" in k-means clustering?
22. How does the "updating step" in k-means clustering work?
23. What is the importance of initialization in the k-means algorithm?
24. Describe some common initialization methods for k-means clustering.
25. What is the role of the sum of squared errors (SSE) in k-means clustering?
26. What are the limitations of the k-means clustering algorithm?
27. How does the "curse of dimensionality" affect the k-means algorithm?
28. What methods can be used to determine the optimal number of clusters in k-means?
29. Explain the concept of "vector quantization" in the context of k-means clustering.
30. What are some applications of k-means clustering in feature learning?
31. How is k-means clustering used in cluster analysis?
32. What is a Gaussian Mixture Model (GMM), and how does it work?
33. How does GMM differ from k-means clustering?
34. What is the role of probabilities in Gaussian Mixture Models?
35. What are the key parameters of a Gaussian Mixture Model?
36. Explain how GMM handles overlapping clusters.
37. In what situations would you prefer to use GMM over k-means?
38. What are the advantages of GMM in comparison to k-means?
39. What is the Expectation-Maximization (EM) method, and how does it work?
40. How is the EM algorithm used to optimize Gaussian Mixture Models?
41. What are the steps involved in the Expectation-Maximization algorithm?
42. What is the difference between the "Expectation" step and the "Maximization" step in EM?
43. How does the EM algorithm handle missing or incomplete data?
44. What are some limitations of the Expectation-Maximization algorithm?
45. What is Principal Component Analysis (PCA), and what is its purpose?
46. How does PCA help in dimensionality reduction?
47. What is the first principal component, and why is it important?
48. How are further principal components calculated after the first one?
49. What is the role of covariance in PCA?
50. Explain the concept of singular value decomposition (SVD) in PCA.
51. What is the difference between eigenvectors and eigenvalues in PCA?
52. How is PCA related to variance in the data?
53. What is the importance of the covariance matrix in PCA?
54. How is PCA used for feature extraction in predictive models?
55. What are the main properties of PCA?
56. What are some typical limitations of PCA in data analysis?
57. Why is it important to center the data before applying PCA?
58. How does PCA perform when the data is non-linear?
59. How is overfitting avoided when using PCA for dimensionality reduction?
60. Describe a real-world application of PCA in predictive modeling.

Unit 5: Parameter Estimation, Model Evaluation and Ensemble Methods

1. What is parameter estimation, and why is it important in statistical modeling?
2. Explain the concept of point estimation in parameter estimation.
3. What is the difference between a point estimator and an interval estimator?
4. What is Maximum Likelihood Estimation (MLE), and how does it work?
5. How is the likelihood function defined in MLE?
6. What are the key assumptions required for applying Maximum Likelihood Estimation?
7. Describe the process of calculating MLE for a given dataset.
8. What is an unbiased estimator?
9. Explain the importance of unbiasedness in point estimation.
10. Give an example of an unbiased estimator in practice.
11. What is the difference between a biased and an unbiased estimator?
12. What are confidence intervals, and how are they used in parameter estimation?
13. How is a confidence interval for one mean constructed?
14. How do you interpret a 95% confidence interval for a population mean?
15. What factors affect the width of a confidence interval for a mean?
16. Explain how to calculate a confidence interval for two means.
17. How is the confidence interval for the difference of two means interpreted?
18. How do you estimate the variance of a population using sample data?
19. What are the steps involved in constructing a confidence interval for variance?
20. What is the relationship between sample size and the accuracy of parameter estimates?
21. Why is model evaluation critical in machine learning?
22. How is ML model validation by humans performed, and what are its limitations?
23. What is the holdout set validation method, and how is it applied in model evaluation?
24. Explain the concept of cross-validation in machine learning.
25. What are the advantages of using cross-validation over the holdout method?
26. How does k-fold cross-validation work?
27. What is the difference between k-fold cross-validation and leave-one-out cross-validation?
28. Explain the leave-one-out cross-validation method and its use cases.
29. What is random subsampling validation, and how is it different from cross-validation?
30. Describe the teach and test method in machine learning validation.
31. What are the strengths and limitations of the teach and test method?
32. How does the bootstrapping method work in ML model validation?
33. What are the advantages of bootstrapping for small datasets?
34. Explain how AI model simulations are used for validation.
35. What is the overriding mechanism method in AI model validation?
36. What is the purpose of using validation sets in machine learning?
37. How do you identify overfitting during model evaluation?
38. What role does cross-validation play in preventing overfitting?
39. How does the ROC curve help in model evaluation?
40. What does the area under the ROC curve (AUC) represent?
41. How is the ROC curve constructed for a binary classifier?
42. What is the relationship between the true positive rate and the false positive rate in an ROC curve?
43. Explain how to interpret a high AUC value in model evaluation.
44. What are ensemble methods in machine learning?

45. What is ensemble theory, and how does it improve model performance?
46. Why does combining multiple models in an ensemble often improve prediction accuracy?
47. How is ensemble size determined in ensemble methods?
48. What is voting-based ensemble learning?
49. How does averaging-based ensemble learning work?
50. What is the key difference between voting and averaging in ensemble methods?
51. Explain the concept of boosting in ensemble learning.
52. How does boosting differ from bagging in ensemble methods?
53. What is the weightage average method in ensemble learning?
54. How does the stacking method work in ensemble learning?
55. What is the role of a meta-learner in stacking?
56. How is bagging used to reduce model variance in ensemble learning?
57. What are the key advantages of bootstrap aggregating (bagging) in machine learning?
58. How is the concept of bootstrapping applied in bagging?
59. What are the differences between bagging and boosting?
60. Describe a real-world application where boosting has significantly improved model performance.