Here's a **hard-level IIT-style quiz** with both **MCQs and Subjective Questions** based on the provided machine learning content.

Machine Learning Quiz (IIT-Level)

Total Marks: 100

Section 1: Multiple Choice Questions (MCQs) – (40 Marks, 2 Marks Each)

1. Which of the following statements about the Bias-Variance tradeoff is true?
a) A high-bias model usually overfits the data.
b) A high-variance model tends to underfit the data.
c) Increasing training data helps in reducing variance.
d) A high-bias model captures too much noise from the data.

2. The Vapnik-Chervonenkis (VC) dimension measures:
a) Model interpretability
b) Model capacity to generalize
c) Number of hidden layers in a neural network
d) Computational complexity of an ML algorithm

3. In Maximum Likelihood Estimation (MLE), the likelihood function represents:
a) The probability of the observed data given the model parameters
b) The probability of model parameters given the observed data
c) The prior probability of the data
d) The posterior probability of the parameters

4. The kernel trick is useful in:
a) Reducing the dimensionality of data
b) Transforming non-linearly separable data into a higher-dimensional space
c) Avoiding overfitting in deep learning models
d) Optimizing loss functions in neural networks

5. The Support Vector Machine (SVM) finds the optimal hyperplane by:
a) Maximizing the within-class variance
b) Minimizing the number of support vectors
c) Maximizing the margin between two classes
d) Reducing computational complexity

6. In Principal Component Analysis (PCA), which of the following is maximized?
a) Within-class variance
b) Between-class variance
c) Total variance of the dataset
d) Entropy of the feature distribution

7. The Expectation-Maximization (EM) algorithm is primarily used for:
a) Supervised learning tasks
b) Clustering and density estimation
c) Regularization in deep learning
d) Hyperparameter tuning

8. In reinforcement learning, the Bellman equation is used to:
a) Minimize the variance of a classifier
b) Optimize kernel functions in SVM
c) Compute the optimal policy for an agent
d) Estimate the probability distribution of data

9. Overfitting in neural networks can be reduced using:
a) Increasing the number of parameters
b) Dropout regularization
c) Decreasing the training data
d) Using only the training dataset for evaluation

10. A common problem in high-dimensional datasets is:
a) Data sparsity
b) Bias-variance tradeoff
c) Model interpretability issues
d) All of the above

11. Which clustering algorithm assumes that data is generated from a mixture of probability distributions?
a) K-Means Clustering
b) DBSCAN
c) Hierarchical Clustering
d) Gaussian Mixture Model (GMM)

12. In Semi-Supervised Learning, labeled data is used to:
a) Train the model entirely
b) Assign labels to all unlabeled data
c) Guide the learning process in combination with unlabeled data
d) Perform reinforcement learning

13. In deep learning, vanishing gradients primarily affect:
a) CNNs
b) Shallow networks
c) RNNs and LSTMs
d) Decision trees

14. What role does Transfer Learning play in deep learning models?
a) Reduces computational cost by reusing pre-trained features
b) Makes models independent of large datasets
c) Eliminates the need for fine-tuning
d) Prevents overfitting entirely

15. Which of the following activation functions is most prone to the vanishing gradient problem?
a) ReLU
b) Leaky ReLU
c) Sigmoid
d) Swish

16. The key difference between RNNs and LSTMs is:
a) LSTMs use gating mechanisms to handle long-term dependencies
b) RNNs have higher computational efficiency than LSTMs
c) LSTMs are only used for text data, while RNNs work for all data types
d) RNNs use attention mechanisms, whereas LSTMs do not

17. The dropout technique in neural networks helps to:
a) Increase overfitting
b) Decrease computation time
c) Regularize the network by preventing co-adaptation of neurons
d) Improve model interpretability

18. The decision boundary of a linear classifier is:
a) Always a straight line or hyperplane
b) Always a curve
c) Can be either a hyperplane or a non-linear curve
d) Unaffected by feature transformation

19. Which metric is best suited for evaluating an imbalanced classification problem?
a) Accuracy
b) Mean Squared Error
c) Precision-Recall AUC
d) Adjusted R-Squared

20. The primary disadvantage of k-Means clustering is:
a) It requires prior knowledge of the number of clusters
b) It cannot handle categorical data
c) It only works for supervised learning
d) It is computationally inefficient for small datasets

Section 2: Subjective Questions (60 Marks, 12 Marks Each)

21. Derive the Maximum Likelihood Estimation (MLE) for a Gaussian distribution and explain its significance in parameter estimation.

22. Show that Linear Discriminant Analysis (LDA) maximizes the ratio of between-class variance to within-class variance and discuss its applications.

23. Explain the kernel trick in Support Vector Machines (SVM) and derive the Lagrangian dual form of the SVM optimization problem.

24. In the context of deep learning, explain how Batch Normalization helps in faster convergence and regularization. Provide mathematical justification.

25. Consider an imbalanced dataset for fraud detection. Discuss the pros and cons of techniques like SMOTE, Cost-sensitive Learning, and Precision-Recall AUC in handling such imbalances.

Total Marks: 100

MCQs: 40 Marks (2 Marks Each) + Subjective: 60 Marks (12 Marks Each)

This quiz is designed at an IIT-level, covering theoretical concepts, mathematical derivations, and real-world applications. Let me know if you need modifications or additional questions! 🚀