Machine Learning Basics

Multiple Choice Questions (1 Mark Each)

1. Which of the following is a supervised learning algorithm?

a) K-Means **b) Decision Tree**

c) DBSCAN d) PCA

2. What does MSE stand for in machine learning?

a) Mean Similarity Estimator **b) Mean Squared Error**

c) Maximum Sampling Estimation d) Multi-Step Evaluation

3. In supervised learning, the dataset is:

a) Unlabeled b) Semi-labeled **c) Labeled** d) Unstructured

4. Which function is used in Logistic Regression to map values between 0 and 1?

a) Tanh b) ReLU **c) Sigmoid** d) Softmax

5. Epoch in machine learning refers to:

a) A single data sample

**b) One complete pass over the entire dataset**

c) A single iteration on a batch

d) The final test phase

6. Overfitting occurs when:  
 a) The model performs well on both training and test data   **b) The model performs well on training data but poorly on test data**   
 c) The model fails to learn the training data  d) The dataset is too small

7. Which of the following best explains reinforcement learning?  
 a) Learning using only labeled data   **b) Learning from reward feedback through trial and error**   
 c) Learning by clustering data points d) Learning using decision trees

8. Batch size affects:  
 a) The number of neurons in the model

**b) The amount of training examples per iteration**  
c) The total dataset size

d) The number of epochs

9. What is the primary goal of unsupervised learning?  
 a) Predict future data   b) Classify labeled data  
 **c) Discover hidden patterns**   d) Validate models

10. In a decision tree, which node represents the final output?  
 a) Root Node   b) Parent Node  
 **c) Leaf Node**   d) Attribute Node

11. Which algorithm is most suitable for spam email detection?  
 a) K-Means   b) Decision Tree  
 **c) Logistic Regression**   d) DBSCAN

12.If a logistic regression model gives a value of 0.7, it predicts:  
 a) Class 0   **b) Class 1**  
 c) Both classes   d) None

13.To prevent overfitting, which technique can be used?  
 **a) Dropout**   b) Increasing bias  
 c) Reducing data   d) Ignoring noise

14.In K-NN, the classification depends on:   
 a) Centroid distance

**b) Majority voting among nearest neighbors**  
c) Decision tree rules

d) Bayesian probability

15.Which clustering technique creates a tree-like structure?  
 a) K-Means   b) DBSCAN  
 **c) Hierarchical**   d) Random Forest

16.What does high variance indicate in a model?  
 **a) Overfitting**   b) Underfitting  
 c) Good generalization   d) Stable prediction

17.Which parameter controls the step size in gradient descent?  
 a) Epoch   b) Batch size  
 **c) Learning rate**   d) Dropout rate

18.Random Forest improves decision tree performance by:  
 a) Removing all outliers   **b) Combining multiple trees**  
 c) Reducing data size   d) Using unsupervised learning

19.In K-Means, centroids are updated until:  
 a) The dataset size changes

**b) There is no change in cluster assignments**  
c) The batch size equals epoch size

d) Gradient descent converges

20.Information Gain is used for:  
 a) Feature scaling

**b) Attribute selection in decision trees**  
c) Model evaluation

d) Overfitting reduction

21.Which model is most prone to underfitting?  
 a) Deep Neural Networks

b) Complex Decision Trees  
**c) Very simple Linear Regression**

d) Random Forest

22.Cross-validation helps in:  
 a) Increasing variance   b) Avoiding bias  
 **c) Evaluating model generalization**   d) Reducing dataset size

23.If a model shows high training accuracy but low test accuracy, it is:  
 a) Underfitting   b) Balanced  
 **c) Overfitting**   d) Optimized

1. Which loss function is commonly used in regression tasks?  
   a) Cross-Entropy   b) Hinge Loss  
   **c) Mean Squared Error**   d) Negative Log-Likelihood

24.In Bayesian statistics, the posterior is:  
 a) The initial belief before data

**b) The updated probability after observing data**  
c) The variance of data

d) The loss function

1. Which scenario best applies reinforcement learning?  
   a) Predicting housing prices   b) Image classification  
   **c) Training a robot to walk**   d) Email spam filtering
2. Which approach would you choose for customer segmentation?  
   a) Logistic Regression   **b) K-Means**  
   c) Random Forest   d) Linear Regression
3. To tune hyperparameters effectively, you would:  
   a) Use training data repeatedly   **b) Use validation sets**  
   c) Ignore test sets   d) Remove epochs

29.Dimensionality reduction is mainly used for:  
 a) Increasing data redundancy

**b) Improving model interpretability**  
c) Enhancing bias

 d) Increasing dropout rate

30.Which method is best for predicting continuous numerical values?  
 a) K-Means   b) Logistic Regression  
 **c) Linear Regression**   d) Naive Bayes

**5 Marks Questions**

1. Define Machine learning?Explain the types of machine learning algorithms?

2.Differentiate between overfitting and underfitting

3.Explain the purpose of using a validation set.

4.Define estimator.Explain the types of estimators.

5.Define Bias and explain its types

6. Define variance and explain its types

7.Explain the Maximum Likelihood estimation in detail.

8.Explain Bayes theorem with an example.

9.Explain the application of Bayesian Statistics.

10.What is supervised learning? Explain the steps involved in supervised learning.

11.Define estimators and explain its types

12.What are the advantages and disadvantages of supervised learning algorithm?

13.Explain the random forests algorithm.

14.Define Regression and explain its types.

15.Explain unsupervised learning in detail with its categories.

16.What is clustering? Explain the k means algorithm.

17.Define hierarchical clustering. Explain Agglomerative clustering with an example.

18.What is hierarchical clustering? Explain Divisive clustering

19.Define Stochastic Gradient Descent (SGD) and explain the loss function with an equation.

20.Explain the steps involved in building a machine learning algorithm.

10 Marks Questions

1.Explain the different types of machine learning: supervised, unsupervised, and reinforcement learning with suitable examples.

2.Discuss hyperparameters in detail

3.What is Bayesian statistics? Explain Bayes’ theorem with applications in machine learning.

4.Explain the K nearset neighbour with an example.

5.Define estimators.Explain the properties of estimators and its types.

6.Explain Decision Tree Classification with attribute selection measures

7.Given the following 2D data points:  
A1(2, 10), A2(2, 5), A3(8, 4), B1(5, 8), B2(7, 5), B3(6, 4), C1(1, 2), C2(4, 9) Apply the K-means algorithm for two iterations to cluster these points into K=3 clusters. Assume initial centroids are randomly chosen as A2(2,5), B2(7,5), and C1(1,2).

8.Given the following data points:  
A(1.0, 1.0), B(1.5, 2.0), C(3.0, 4.0), D(5.0, 7.0), E(3.5, 5.0), F(4.5, 5.0), G(3.5, 4.5)And initial cluster centroids: C1 = (1.0, 1.0) C2 = (5.0, 7.0).Apply the K-means algorithm for two iterations to cluster these data points into two clusters (K=2).

9.Perform Agglomerative Hierarchical Clustering on the following dataset of points using Euclidean distance as the similarity measure:A (1, 1),B (2, 3),C (3, 5),D (4, 5),E (6, 6),F (7, 5)

a) Construct the distance matrix for all pairs of points.  
b) Show the step-by-step merging process until all points are clustered into one group.  
c) Draw the dendrogram to represent the clustering hierarchy.

10.Explain the working of Stochastic Gradient Descent (SGD) .

11.Explain the purpose of using a validation set with an example

12.Explain the challenges in machine learning and how deep learning overcomes them.

13.Explain the steps involved in building a machine learning algorithm with example code.