socrative
A showbie PRODUCT

Name _____
Date _____

DA -04

Score _____

1. Which of the following are model parameters, not hyperparameters?

(A) Weights in Logistic Regression
(B) Regularization strength C in Logistic Regression
(C) max_depth in Decision Tree
(D) Support vectors in SVM

2. In a Decision Tree classifier, which hyperparameter prevents overfitting by limiting tree complexity?

(A) max_depth
(B) min_samples_split
(C) min_samples_leaf
(D) All of the above

3. When tuning SVM with RBF kernel, what is the effect of increasing gamma?

(A) Makes decision boundary smoother (less overfitting)
(B) Makes decision boundary more complex (risk of overfitting)
(C) Increases margin width
(D) Has no effect

4. Which hyperparameter in Logistic Regression controls the amount of regularization?

(A) penalty
(B) solver
(C) C
(D) max_iter

5. You run a GridSearchCV with a very large parameter grid. Which of the following is MOST likely?

(A) Fast search but poor performance

(B) Very slow search but exhaustive

(C) Random exploration of parameter space

(D) Tuning may fail due to no refitting

6. Which statement about GridSearchCV is FALSE?

(A) It can use multiple scoring metrics

(B) It can perform cross-validation

(C) It automatically chooses hyperparameters during training updates

(D) It can refit the best model

7. Which hyperparameter in SVM controls the trade-off between margin size and misclassification errors?

(A) gamma

(B) kernel

(C) C

(D) degree

8. You use GridSearchCV with cv=5 on a dataset of 1000 samples. How many models are trained per hyperparameter combination?

(A) 5

(B) 1000

(C) 200

(D) 1

9. Which advanced method discards poorly performing hyperparameter configurations early to save time?

(A) Grid Search

(B) Random Search

(C) Hyperband / Successive Halving

(D) Cross-validation

10. If you include the test set during hyperparameter tuning, what problem occurs?

(A) Data Leakage

(B) Over-regularization

(C) Underfitting

(D) Faster convergence

11. You are tuning Logistic Regression for the Breast Cancer Wisconsin dataset. You try values of C = $[0.01, 0.1, 1, 10]$. When C= 0.01, the model underfits (low train and test accuracy). What is the correct reason?

(A) C=0.01 applies strong regularization, shrinking coefficients too much.

(B) C=0.01 applies weak regularization, allowing overfitting.

(C) Logistic regression cannot handle this dataset.

(D) C has no effect on model complexity.

12. You train a Decision Tree without setting max_depth, and it achieves 100% training accuracy but only 70% test accuracy. What hyperparameter should you tune first to reduce overfitting?

(A) criterion

(B) max_depth

(C) splitter

(D) random_state

13. On the Breast Cancer dataset, you use an SVM with RBF kernel. With gamma= 0.0001, the model performs poorly (both train and test accuracy are low). What is the most likely reason?

(A) Gamma is too low, decision boundary is too simple.

(B) Gamma is too high, causing overfitting.

(C) Kernel is wrong; should use linear kernel.

(D) C value is too high.

14. You perform GridSearchCV with the following parameter grid for Decision Tree: param_grid = {"max_depth": [3, 5, 7, None], "min_samples_split": [2, 5, 10]} If cv=5 and your dataset has 1000 samples, how many models are trained in total?

(A) 12

(B) 60

(C) 20

(D) 15

15. While tuning hyperparameters for Logistic Regression, a student mistakenly used the test set inside GridSearchCV. Which of the following is MOST likely to happen?

(A) Test accuracy will be underestimated.

(B) Model will overfit training data only.

(C) Hyperparameters will be biased toward test set, giving overly optimistic results.

(D) Nothing significant, since CV already prevents overfitting.

16. For an SVM with RBF kernel, you want to tune C and gamma. The search space is very large. Which tuning method is most efficient to try first?

(A) Grid Search

(B) Random Search

(C) Manual search

(D) Successive Halving

17. You run GridSearchCV with Logistic Regression and find that the best C= 1.0 . How do you get the best model to use for predictions?

(A) grid.best_params_

(B) grid.best_score_

(C) grid.best_estimator_

(D) grid.cv_results_

18. You restrict max_depth=2 for a Decision Tree on Breast Cancer dataset, but accuracy is only ~65% on both train and test sets. What is happening?

(A) Overfitting

(B) Underfitting

(C) Data Leakage

(D) High variance

19. Which hyperparameter in SVM directly controls how much misclassification is tolerated?

(A) gamma

(B) degree

(C) kernel

(D) C

20. Which of the following best describes ensemble learning?

(A) Training a single model with more epochs to improve accuracy

(B) Combining multiple models to achieve better performance

(C) Using deep learning models instead of shallow ones

(D) Training models only in parallel on different machines

21. What is the main idea behind bagging?

(A) Assigning higher weights to misclassified examples

(B) Combining weak learners sequentially

(C) Training base learners on different random subsets of data with replacement

(D) Reducing bias by averaging predictions from one model

22. In boosting, each new learner is trained to:

(A) Reduce correlation between models

(B) Correct mistakes made by previous learners

(C) Train on a completely random subset of data

(D) Maximize variance of predictions

23. Which of the following ensemble methods is an example of parallel ensemble learning?

(A) AdaBoost

(B) Gradient Boosting

(C) Random Forest

(D) XGBoost

24. What type of sampling is used in bagging?

(A) Random sampling without replacement

(B) Random sampling with replacement (bootstrap sampling)

(C) Stratified sampling

(D) Systematic sampling

25. In Random Forest, why is feature randomness (max_features) used?

(A) To reduce the training time of each tree

(B) To ensure each tree is different and less correlated

(C) To increase the number of leaf nodes

(D) To decrease the variance of each tree individually

26. Which of the following is a major disadvantage of ensemble learning?

(A) It always increases bias

(B) It is less interpretable and computationally more expensive

(C) It cannot handle non-linear data

(D) It does not reduce variance

27. In boosting, if a data point is misclassified repeatedly, what usually happens?

(A) Its weight decreases in the next iteration

(B) Its weight increases in the next iteration

(C) It gets removed from the dataset

(D) The model ignores it

28. Which of the following pairs correctly matches method with base idea?

(A) Bagging – Reduce bias

(B) Boosting – Reduce variance

(C) Bagging – Reduce variance

(D) Boosting – Reduce interpretability only

29. You are training a model for a medical diagnosis dataset (high-stakes, small dataset). The base decision tree tends to overfit. Which ensemble method would be more suitable?

(A) Bagging

(B) Boosting

(C) Logistic Regression

(D) K-Means

30. In clustering, the goal is to

(A) Minimize within-cluster similarity

(B) Maximize between-cluster similarity

(C) Both (a) and (b)

(D) None of the above

31. K-Means clustering requires

(A) Pre-defined number of clusters (k)

(B) Distance metric

(C) Random initialization of centroids

(D) All of the above

32. The main drawback of K-Means clustering is

(A) It can only handle categorical data

(B) It requires knowing the number of clusters in advance

(C) It is not suitable for large datasets

(D) It doesn't use distance metrics

33. Which metric is commonly used to evaluate clustering results without ground truth labels

(A) Accuracy

(B) Adjusted Rand Index

(C) Silhouette Score

(D) F1 Score

34. Dimensionality reduction is mainly used to

(A) Increase the number of features

(B) Reduce noise and redundancy in data

(C) Increase model complexity

(D) Improve overfitting

35. In PCA, the new axes (principal components) are

(A) Original features scaled

(B) Orthogonal linear combinations of features

(C) Random projections of features

(D) Nonlinear transformations

36. The first principal component (PC1) captures

(A) The smallest variance in the data

(B) The largest variance in the data

(C) Only categorical variance

(D) No variance at all

37. Eigenvalues in PCA represent

(A) Variance explained by each principal component

(B) The angle between components

(C) The mean of the dataset

(D) The correlation coefficient

38. Which step comes first in PCA

(A) Compute eigenvalues/eigenvectors

(B) Standardize the dataset

(C) Project data onto new space

(D) Compute covariance matrix

39. If two variables are highly correlated, PCA will

(A) Drop one variable

(B) Merge them into a principal component

(C) Ignore both variables

(D) Always assign them to PC2

40. The curse of dimensionality generally refers to

(A) Increase in computation cost as data grows in rows

(B) Problems when data has too many features

(C) Errors in model labeling

(D) Lack of training samples