Advanced Data Science Principles

Data Science · Answer Key · 15 Questions

1. In the context of the Bias-Variance Tradeoff, what does the 'irreducible error' represent?

A) The error caused by high-bias models
B) The variance component of the mean squared error
C) The noise term in the true data-generating process
D) The bias introduced by feature selection

2. Which theorem states that for any optimization algorithm, every representation of all possible problems will have the same performance when averaged over all possible problems?

A) The No Free Lunch Theorem
B) The Central Limit Theorem
C) The Gauss-Markov Theorem
D) The Stone-Weierstrass Theorem

3. In Information Theory, what is the mathematical definition of Kullback-Leibler (KL) divergence?

A) A symmetric measure of distance between two probability distributions
B) The entropy of a joint distribution minus the entropy of marginals
C) A measure of how one probability distribution differs from a second, reference distribution
D) The log-likelihood ratio of two nested models

4. Which regularization technique adds a penalty equal to the sum of the absolute values of the coefficients?

A) Ridge Regression (L2)
B) Lasso Regression (L1)
C) Elastic Net
D) Elastic Dropout

5. In the context of Support Vector Machines (SVM), what is the function of the 'kernel trick'?

A) To normalize input data to a mean of zero
B) To map input features into a higher-dimensional space to allow linear separation
C) To reduce the number of support vectors for faster computation
D) To optimize the selection of the hyperparameter C

6. What is the primary condition required for the Gauss-Markov Theorem to hold regarding Ordinary Least Squares (OLS) estimators?

A) Errors must follow a non-parametric distribution
B) Errors must be homoscedastic and uncorrelated with mean zero
C) The number of predictors must exceed the number of observations
D) The model must include an interaction term

7. In Gradient Boosting, what does the 'learning rate' (or shrinkage) parameter fundamentally control?

A) The depth of the individual decision trees
B) The contribution of each successive tree to the final prediction
C) The threshold for pruning leaf nodes
D) The size of the initial bootstrap sample

8. Which property of a stationary time series implies that the joint distribution of the series is invariant to time shifts?

A) Weak stationarity
B) Strict stationarity
C) Cyclostationarity
D) Non-ergodicity

9. What is the 'Curse of Dimensionality' in the context of Euclidean distance?

A) The computational complexity grows exponentially with the number of samples
B) The ratio of the distance to the nearest and farthest neighbor tends toward one in high dimensions
C) The model becomes increasingly biased as dimensions increase
D) The variance of the model decreases as features are added

10. In Bayesian inference, what is the term for the distribution representing prior beliefs updated with observed data?

A) Likelihood function
B) Marginal distribution
C) Posterior distribution
D) Conjugate prior

11. Which activation function is defined as f(x) = ln(1 + e^x) and is known as the smooth approximation of the ReLU function?

A) Sigmoid
B) Softplus
C) Tanh
D) Leaky ReLU

12. In cluster analysis, what is the objective function minimized by the K-Means algorithm?

A) Within-cluster sum of squares (WCSS)
B) Between-cluster variance
C) The Silhouette coefficient
D) The Calinski-Harabasz index

13. Which statistical method is specifically designed to handle multicollinearity by transforming variables into uncorrelated components?

A) Principal Component Regression (PCR)
B) Stepwise Regression
C) Ordinary Least Squares
D) Lasso Regression

14. What does the Vapnik-Chervonenkis (VC) dimension measure in statistical learning theory?

A) The computational speed of an algorithm
B) The capacity (complexity) of a statistical classification algorithm
C) The minimum number of samples needed to achieve zero bias
D) The variance of an estimator in high dimensions

15. In a hidden Markov model, what does the Viterbi algorithm compute?

A) The probability of a specific observation sequence
B) The most likely sequence of hidden states given an observation sequence
C) The parameters that maximize the probability of the observations
D) The stationary distribution of the Markov chain