Multiple Choice Questions

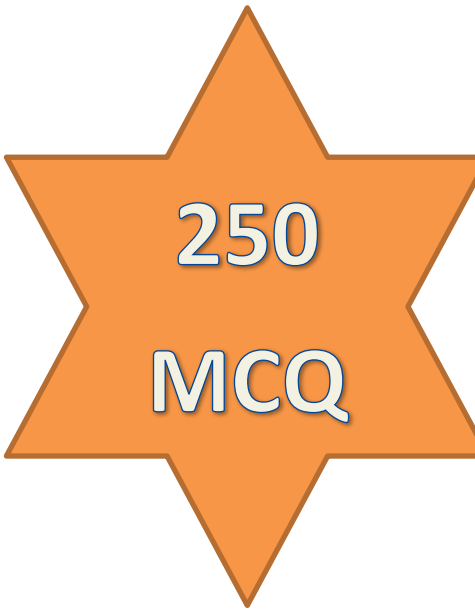
On

Machine Learning

University Academy
Teaching|Training|Informative

By

Mr. Sandeep Vishwakarma
Assistant Professor
Dr. A.P.J. Abdul Kalam
Technical University,Lucknow



MACHINE LEARNING

DETAILED SYLLABUS

UNIT-I

INTRODUCTION – Well defined learning problems, Designing a Learning System, Issues in Machine Learning; THE CONCEPT LEARNING TASK - General-to-specific ordering of hypotheses, Find-S, List then eliminate algorithm, Candidate elimination algorithm, Inductive bias

UNIT-II

DECISION TREE LEARNING - Decision tree learning algorithm-Inductive bias- Issues in Decision tree learning;
ARTIFICIAL NEURAL NETWORKS – Perceptrons, Gradient descent and the Delta rule, Adaline,
Multilayer networks, Derivation of backpropagation rule Backpropagation AlgorithmConvergence,
Generalization;

UNIT-III

Evaluating Hypotheses: Estimating Hypotheses Accuracy, Basics of sampling Theory, Comparing
Learning Algorithms;
Bayesian Learning: Bayes theorem, Concept learning, Bayes Optimal Classifier, Naïve Bayes classifier, Bayesian belief networks, EM algorithm;

UNIT-IV

Computational Learning Theory: Sample Complexity for Finite Hypothesis spaces, Sample Complexity for Infinite Hypothesis spaces, The Mistake Bound Model of Learning;
INSTANCE-BASED LEARNING – k-Nearest Neighbour Learning, Locally Weighted Regression,
Radial basis function networks, Case-based learning

UNIT-V

Genetic Algorithms: an illustrative example, Hypothesis space search, Genetic Programming, Models of Evolution and Learning; Learning first order rules-sequential covering algorithms-General to specific beam search-FOIL; REINFORCEMENT LEARNING - The Learning Task, Q Learning.

1. What is Machine Learning (ML)?
(A) The autonomous acquisition of knowledge through the use of manual programs

(B) The selective acquisition of knowledge through the use of computer programs
(C) The selective acquisition of knowledge through the use of manual programs
(D) The autonomous acquisition of knowledge through the use of computer programs
Answer
Correct option is D

2. Father of Machine Learning (ML)
(A) Geoffrey Chaucer
(B) Geoffrey Hill
(C) Geoffrey Everest Hinton
(D) None of the above
Answer
Correct option is C

3. Which is FALSE regarding regression?
(A) It may be used for interpretation
(B) It is used for prediction
(C) It discovers causal relationships
(D) It relates inputs to outputs
Answer
Correct option is C

4. Choose the correct option regarding machine learning (ML) and artificial intelligence (AI)
(A) ML is a set of techniques that turns a dataset into a software
(B) AI is a software that can emulate the human mind
(C) ML is an alternate way of programming intelligent machines
(D) All of the above
Answer
Correct option is D

5. Which of the factors affect the performance of the learner system does not include?
(A) Good data structures
(B) Representation scheme used
(C) Training scenario
(D) Type of feedback
Answer
Correct option is A

6. In general, to have a well-defined learning problem, we must identity which of the following
(A) The class of tasks
(B) The measure of performance to be improved
(C) The source of experience
(D) All of the above
Answer
Correct option is D

7. Successful applications of ML
(A) Learning to recognize spoken words
(B) Learning to drive an autonomous vehicle
(C) Learning to classify new astronomical structures
(D) Learning to play world-class backgammon
(E) All of the above
Answer
Correct option is E

8. Which of the following does not include different learning methods
(A) Analogy
(B) Introduction
(C) Memorization
(D) Deduction
Answer
Correct option is B

9. In language understanding, the levels of knowledge that does not include?
(A) Empirical
(B) Logical
(C) Phonological
(D) Syntactic
Answer
Correct option is A

10. Designing a machine learning approach involves:-
(A) Choosing the type of training experience
(B) Choosing the target function to be learned
(C) Choosing a representation for the target function
(D) Choosing a function approximation algorithm
(E) All of the above
Answer
Correct option is E

11. Concept learning inferred a _____ valued function from training examples of its input and output.
(A) Decimal
(B) Hexadecimal
(C) Boolean
(D) All of the above
Answer
Correct option is C

12. Which of the following is not a supervised learning?
(A) Naive Bayesian
(B) PCA
(C) Linear Regression
(D) Decision Tree
Answer
Correct option is B

13. What is Machine Learning?
(i) Artificial Intelligence
(ii) Deep Learning
(iii) Data Statistics
(A) Only (i)
(B) (i) and (ii)
(C) All
(D) None
Answer
Correct option is B

14. What kind of learning algorithm for "Facial identities or facial expressions"?
(A) Prediction
(B) Recognition Patterns
(C) Generating Patterns
(D) Recognizing Anomalies
Answer
Correct option is B

15. Which of the following is not type of learning?
(A) Unsupervised Learning
(B) Supervised Learning
(C) Semi-unsupervised Learning
(D) Reinforcement Learning
Answer
Correct option is C

16. Real-Time decisions, Game AI, Learning Tasks, Skill Aquisition, and Robot Navigation are applications of which of the folowing
(A) Supervised Learning: Classification
(B) Reinforcement Learning
(C) Unsupervised Learning: Clustering
(D) Unsupervised Learning: Regression
Answer
Correct option is B

17. Targetted marketing, Recommended Systems, and Customer Segmentation are applications in which of the following
(A) Supervised Learning: Classification
(B) Unsupervised Learning: Clustering
(C) Unsupervised Learning: Regression
(D) Reinforcement Learning
Answer
Correct option is B

18. Fraud Detection, Image Classification, Diagnostic, and Customer Retention are applications in which of the following
(A) Unsupervised Learning: Regression
(B) Supervised Learning: Classification

(C) Unsupervised Learning: Clustering
(D) Reinforcement Learning
Answer
Correct option is B

19. Which of the following is not function of symbolic in the various function representation of Machine Learning?
(A) Rules in propotional Logic
(B) Hidden-Markov Models (HMM)
(C) Rules in first-order predicate logic
(D) Decision Trees
Answer
Correct option is B

20. Which of the following is not numerical functions in the various function representation of Machine Learning?
(A) Neural Network
(B) Support Vector Machines
(C) Case-based
(D) Linear Regression
Answer
Correct option is C

21. FIND-S Algorithm starts from the most specific hypothesis and generalize it by considering only _____ examples.
(A) Negative
(B) Positive
(C) Negative or Positive
(D) None of the above
Answer
Correct option is B

22. FIND-S algorithm ignores _____ examples.
(A) Negative
(B) Positive
(C) Both
(D) None of the above
Answer
Correct option is A

23. The Candidate-Elimination Algorithm represents the _____.
(A) Solution Space
(B) Version Space
(C) Elimination Space
(D) All of the above
Answer
Correct option is B

24. Inductive learning is based on the knowledge that if something happens a lot it is likely to be generally.

(A) True
(B) False
Answer
Correct option is A

25. Inductive learning takes examples and generalizes rather than starting with _____ knowledge.
(A) Inductive
(B) Existing
(C) Deductive
(D) None of these
Answer
Correct option is B

26. A drawback of the FIND-S is that, it assumes the consistency within the training set.
(A) True
(B) False
Answer
Correct option is A

27. What strategies can help reduce overfitting in decision trees?
(i) Enforce a maximum depth for the tree
(ii) Enforce a minimum number of samples in leaf nodes
(iii) Pruning
(iv) Make sure each leaf node is one pure class
(A) All
(B) (i), (ii) and (iii)
(C) (i), (iii), (iv)
(D) None
Answer
Correct option is B

28. Which of the following is a widely used and effective machine learning algorithm based on the idea of bagging?
(A) Decision Tree
(B) Random Forest
(C) Regression
(D) Classification
Answer
Correct option is B

29. To find the minimum or the maximum of a function, we set the gradient to zero because which of the following
(A) Depends on the type of problem
(B) The value of the gradient at extrema of a function is always zero
(C) Both (A) and (B)
(D) None of these
Answer
Correct option is B

30. Which of the following is a disadvantage of decision trees?
(A) Decision trees are prone to be overfit
(B) Decision trees are robust to outliers
(C) Factor analysis
(D) None of the above
Answer
Correct option is A

31. What is perceptron?
(A) A single layer feed-forward neural network with pre-processing
(B) A neural network that contains feedback
(C) A double layer auto-associative neural network
(D) An auto-associative neural network
Answer
Correct option is A

32. Which of the following is true for neural networks?
(i) The training time depends on the size of the network.
(ii) Neural networks can be simulated on a conventional computer.
(iii) Artificial neurons are identical in operation to biological ones.
(A) All
(B) Only (ii)
(C) (i) and (ii)
(D) None
Answer
Correct option is C

33. What are the advantages of neural networks over conventional computers?
(i) They have the ability to learn by example.
(ii) They are more fault tolerant.
(iii)They are more suited for real time operation due to their high 'computational' rates.
(A) (i) and (ii)
(B) (i) and (iii)
(C) Only (i)
(D) All
(E) None
Answer
Correct option is D

34. What is Neuro software?
(A) It is software used by Neurosurgeon
(B) Designed to aid experts in real world
(C) It is powerful and easy neural network
(D) A software used to analyze neurons
Answer
Correct option is C

35. Which is true for neural networks?
(A) Each node computes it's weighted input

(B) Node could be in excited state or non-excited state
(C) It has set of nodes and connections
(D) All of the above
Answer
Correct option is D

36. What is the objective of backpropagation algorithm?
(A) To develop learning algorithm for multilayer feedforward neural network, so that network can be trained to capture the mapping implicitly
(B) To develop learning algorithm for multilayer feedforward neural network
(C) To develop learning algorithm for single layer feedforward neural network
(D) All of the above
Answer
Correct option is A

37. Which of the following is true?
Single layer associative neural networks do not have the ability to:-
(i) Perform pattern recognition
(ii) Find the parity of a picture
(iii) Determine whether two or more shapes in a picture are connected or not
(A) (ii) and (iii)
(B) Only (ii)
(C) All
(D) None
Answer
Correct option is A

38. The backpropagation law is also known as generalized delta rule.
(A) True
(B) False
Answer
Correct option is A

38. Which of the following is true?
(i) On average, neural networks have higher computational rates than conventional computers.
(ii) Neural networks learn by example.
(iii) Neural networks mimic the way the human brain works.
(A) All
(B) (ii) and (iii)
(C) (i), (ii) and (iii)
(D) None
Answer
Correct option is A

39. What is true regarding backpropagation rule?
(A) Error in output is propagated backwards only to determine weight updates
(B) There is no feedback of signal at nay stage

(C) It is also called generalized delta rule
(D) All of the above
Answer
Correct option is D

40. There is feedback in final stage of backpropagation algorithm.
(A) True
(B) False
Answer
Correct option is B

41. An auto-associative network is
(A) A neural network that has only one loop
(B) A neural network that contains feedback
(C) A single layer feed-forward neural network with pre-processing
(D) A neural network that contains no loops
Answer
Correct option is B

42. A 3-input neuron has weights 1, 4 and 3. The transfer function is linear with the constant of proportionality being equal to 3. The inputs are 4, 8 and 5 respectively. What will be the output?
(A) 139
(B) 153
(C) 612
(D) 160
Answer
Correct option is B

43. What of the following is true regarding backpropagation rule?
(A) Hidden layers output is not all important, they are only meant for supporting input and output layers
(B) Actual output is determined by computing the outputs of units for each hidden layer
(C) It is a feedback neural network
(D) None of the above
Answer
Correct option is B

44. What is back propagation?
(A) It is another name given to the curvy function in the perceptron
(B) It is the transmission of error back through the network to allow weights to be adjusted so that the network can learn
(C) It is another name given to the curvy function in the perceptron
(D) None of the above
Answer
Correct option is B

45. The general limitations of back propagation rule is/are
(A) Scaling

(B) Slow convergence
(C) Local minima problem
(D) All of the above
Answer
Correct option is D

46. What is the meaning of generalized in statement "backpropagation is a generalized delta rule" ?
(A) Because delta is applied to only input and output layers, thus making it more simple and generalized
(B) It has no significance
(C) Because delta rule can be extended to hidden layer units
(D) None of the above
Answer
Correct option is C

47. Neural Networks are complex _____ functions with many parameters.
(A) Linear
(B) Non linear
(C) Discreate
(D) Exponential
Answer
Correct option is A

48. The general tasks that are performed with backpropagation algorithm
(A) Pattern mapping
(B) Prediction
(C) Function approximation
(D) All of the above
Answer
Correct option is D

49. Backpropagaion learning is based on the gradient descent along error surface.
(A) True
(B) False
Answer
Correct option is A

50. In backpropagation rule, how to stop the learning process?
(A) No heuristic criteria exist
(B) On basis of average gradient value
(C) There is convergence involved
(D) None of these
Answer
Correct option is B

51. Applications of NN (Neural Network)
(A) Risk management
(B) Data validation

(C) Sales forecasting
(D) All of the above
Answer
Correct option is D

52. The network that involves backward links from output to the input and hidden layers is known as
(A) Recurrent neural network
(B) Self organizing maps
(C) Perceptrons
(D) Single layered perceptron
Answer
Correct option is A

53. Decision Tree is a display of an algorithm.
(A) True
(B) False
Answer
Correct option is A

54. Which of the following is/are the decision tree nodes?
(A) End Nodes
(B) Decision Nodes
(C) Chance Nodes
(D) All of the above
Answer
Correct option is D

55. End Nodes are represented by which of the following
(A) Solar street light
(B) Triangles
(C) Circles
(D) Squares
Answer
Correct option is B

56. Decision Nodes are represented by which of the following
(A) Solar street light
(B) Triangles
(C) Circles
(D) Squares
Answer
Correct option is D

57. Chance Nodes are represented by which of the following
(A) Solar street light
(B) Triangles
(C) Circles
(D) Squares
Answer

Correct option is C

58. Advantage of Decision Trees
(A) Possible Scenarios can be added
(B) Use a white box model, if given result is provided by a model
(C) Worst, best and expected values can be determined for different scenarios
(D) All of the above
Answer
Correct option is D

59. _____ terms are required for building a bayes model.
(A) 1
(B) 2
(C) 3
(D) 4
Answer
Correct option is C

60. Which of the following is the consequence between a node and its predecessors while creating bayesian network?
(A) Conditionally independent
(B) Functionally dependent
(C) Both Conditionally dependant & Dependant
(D) Dependent
Answer
Correct option is A

61. Why it is needed to make probabilistic systems feasible in the world?
(A) Feasibility
(B) Reliability
(C) Crucial robustness
(D) None of the above
Answer
Correct option is C

62. Bayes rule can be used for:-
(A) Solving queries
(B) Increasing complexity
(C) Answering probabilistic query
(D) Decreasing complexity
Answer
Correct option is C

63. _____ provides way and means of weighing up the desirability of goals and the likelihood of achieving them.
(A) Utility theory
(B) Decision theory
(C) Bayesian networks
(D) Probability theory
Answer

Correct option is A

64. Which of the following provided by the Bayesian Network?
(A) Complete description of the problem
(B) Partial description of the domain
(C) Complete description of the domain
(D) All of the above
Answer
Correct option is C

65. Probability provides a way of summarizing the _____ that comes from our laziness and ignorances.
(A) Belief
(B) Uncertaintity
(C) Joint probability distributions
(D) Randomness
Answer
Correct option is B

66. The entries in the full joint probability distribution can be calculated as
(A) Using variables
(B) Both Using variables & information
(C) Using information
(D) All of the above
Answer
Correct option is C

67. Causal chain (For example, Smoking cause cancer) gives rise to:-
(A) Conditionally Independence
(B) Conditionally Dependence
(C) Both
(D) None of the above
Answer
Correct option is A

68. The bayesian network can be used to answer any query by using:-
(A) Full distribution
(B) Joint distribution
(C) Partial distribution
(D) All of the above
Answer
Correct option is B

69. Bayesian networks allow compact specification of:-
(A) Joint probability distributions
(B) Belief
(C) Propositional logic statements
(D) All of the above
Answer
Correct option is A

70. The compactness of the bayesian network can be described by
(A) Fully structured
(B) Locally structured
(C) Partially structured
(D) All of the above
Answer
Correct option is B

71. The Expectation Maximization Algorithm has been used to identify conserved domains in unaligned proteins only. State True or False.
(A) True
(B) False
Answer
Correct option is B

72. Which of the following is correct about the Naive Bayes?
(A) Assumes that all the features in a dataset are independent
(B) Assumes that all the features in a dataset are equally important
(C) Both
(D) All of the above
Answer
Correct option is C

73. Which of the following is false regarding EM Algorithm?
(A) The alignment provides an estimate of the base or amino acid composition of each column in the site
(B) The column-by-column composition of the site already available is used to estimate the probability of finding the site at any position in each of the sequences
(C) The row-by-column composition of the site already available is used to estimate the probability
(D) None of the above
Answer
Correct option is C

74. Naïve Bayes Algorithm is a _____ learning algorithm.
(A) Supervised
(B) Reinforcement
(C) Unsupervised
(D) None of these
Answer
Correct option is A

75. EM algorithm includes two repeated steps, here the step 2 is _____.
(A) The normalization
(B) The maximization step
(C) The minimization step
(D) None of the above
Answer
Correct option is C

76. Examples of Naïve Bayes Algorithm is/are
(A) Spam filtration
(B) Sentimental analysis
(C) Classifying articles
(D) All of the above
Answer
Correct option is D

77. In the intermediate steps of "EM Algorithm", the number of each base in each column is determined and then converted to fractions.
(A) True
(B) False
Answer
Correct option is A

78. Naïve Bayes algorithm is based on _____ and used for solving classification problems.
(A) Bayes Theorem
(B) Candidate elimination algorithm
(C) EM algorithm
(D) None of the above
Answer
Correct option is A

79. Types of Naïve Bayes Model:
(A) Gaussian
(B) Multinomial
(C) Bernoulli
(D) All of the above
Answer
Correct option is D

80. Disadvantages of Naïve Bayes Classifier:
(A) Naive Bayes assumes that all features are independent or unrelated, so it cannot learn the relationship between features.
(B) It performs well in Multi-class predictions as compared to the other Algorithms.
(C) Naïve Bayes is one of the fast and easy ML algorithms to predict a class of datasets.
(D) It is the most popular choice for text classification problems.
Answer
Correct option is A

81. The benefit of Naïve Bayes:-
(A) Naïve Bayes is one of the fast and easy ML algorithms to predict a class of datasets.
(B) It is the most popular choice for text classification problems.
(C) It can be used for Binary as well as Multi-class Classifications.
(D) All of the above
Answer
Correct option is D

82. In which of the following types of sampling the information is carried out under the opinion of an expert?
(A) Convenience sampling
(B) Judgement sampling
(C) Quota sampling
(D) Purposive sampling
Answer
Correct option is B

83. Full form of MDL.
(A) Minimum Description Length
(B) Maximum Description Length
(C) Minimum Domain Length
(D) None of these
Answer
Correct option is A

84. For the analysis of ML algorithms, we need
(A) Computational learning theory
(B) Statistical learning theory
(C) Both A & B
(D) None of these
Answer
Correct option is C

85. PAC stand for
(A) Probably Approximate Correct
(B) Probably Approx Correct
(C) Probably Approximate Computation
(D) Probably Approx Computation
Answer
Correct option is A

86. _____ of hypothesis h with respect to target concept c and distribution D , is the probability that h will misclassify an instance drawn at random according to D
(A) True Error
(B) Type 1 Error
(C) Type 2 Error
(D) None of these
Answer
Correct option is A

87. Statement: True error defined over entire instance space, not just training data
(A) True
(B) False
Answer

Correct option is A

88. What are the area CLT comprised of?
(A) Sample Complexity
(B) Computational Complexity
(C) Mistake Bound
(D) All of these
Answer
Correct option is D

88. What area of CLT tells "How many examples we need to find a good hypothesis ?"?
(A) Sample Complexity
(B) Computational Complexity
(C) Mistake Bound
(D) None of these
Answer
Correct option is A

89. What area of CLT tells "How much computational power we need to find a good hypothesis ?"?
(A) Sample Complexity
(B) Computational Complexity
(C) Mistake Bound
(D) None of these
Answer
Correct option is B

90. What area of CLT tells "How many mistakes we will make before finding a good hypothesis ?"?
(A) Sample Complexity
(B) Computational Complexity
(C) Mistake Bound
(D) None of these
Answer
Correct option is C

91. (For question no. 9 and 10) Can we say that concept described by conjunctions of Boolean literals are PAC learnable?
(A) Yes
(B) No
Answer
Correct option is A

92. How large is the hypothesis space when we have n Boolean attributes?
(A) $|H| = 3^n$
(B) $|H| = 2^n$
(C) $|H| = 1^n$
(D) $|H| = 4^n$
Answer
Correct option is A

93. The VC dimension of hypothesis space H1 is larger than the VC dimension of hypothesis space H2. Which of the following can be inferred from this?
(A) The number of examples required for learning a hypothesis in H1 is larger than the number of examples required for H2
(B) The number of examples required for learning a hypothesis in H1 is smaller than the number of examples required for H2.
(C) No relation to number of samples required for PAC learning.
Answer
Correct option is A

94. For a particular learning task, if the requirement of error parameter changes from 0.1 to 0.01. How many more samples will be required for PAC learning?
(A) Same
(B) 2 times
(C) 1000 times
(D) 10 times
Answer
Correct option is D

95. Computational complexity of classes of learning problems depends on which of the following?
(A) The size or complexity of the hypothesis space considered by learner
(B) The accuracy to which the target concept must be approximated
(C) The probability that the learner will output a successful hypothesis
(D) All of these
Answer
Correct option is D

96. The instance-based learner is a _____
(A) Lazy-learner
(B) Eager learner
(C) Can't say
Answer
Correct option is A

97. When to consider nearest neighbour algorithms?
(A) Instance map to point in k^n
(B) Not more than 20 attributes per instance
(C) Lots of training data
(D) None of these
(E) A, B & C
Answer
Correct option is E

98. What are the advantages of Nearest neighbour alogo?
(A) Training is very fast
(B) Can learn complex target functions
(C) Don't lose information
(D) All of these
Answer

Correct option is D

99. What are the difficulties with k-nearest neighbour algo?
(A) Calculate the distance of the test case from all training cases
(B) Curse of dimensionality
(C) Both A & B
(D) None of these
Answer
Correct option is C

100. What if the target function is real valued in kNN algo?
(A) Calculate the mean of the k nearest neighbours
(B) Calculate the SD of the k nearest neighbour
(C) None of these
Answer
Correct option is A

101. What is/are true about Distance-weighted KNN?
(A) The weight of the neighbour is considered
(B) The distance of the neighbour is considered
(C) Both A & B
(D) None of these
Answer
Correct option is C

102. What is/are advantage(s) of Distance-weighted k-NN over k-NN?
(A) Robust to noisy training data
(B) Quite effective when a sufficient large set of training data is provided
(C) Both A & B
(D) None of these
Answer
Correct option is C

103. What is/are advantage(s) of Locally Weighted Regression?
(A) Pointwise approximation of complex target function
(B) Earlier data has no influence on the new ones
(C) Both A & B
(D) None of these
Answer
Correct option is C

104. The quality of the result depends on (LWR)
(A) Choice of the function
(B) Choice of the kernel function K
(C) Choice of the hypothesis space H
(D) All of these
Answer
Correct option is D

105. How many types of layer in radial basis function neural networks?
(A) 3
(B) 2
(C) 1
(D) 4
Answer
Correct option is A, Input layer, Hidden layer, and Output layer

106. The neurons in the hidden layer contains Gaussian transfer function whose output are _____ to the distance from the centre of the neuron.
(A) Directly
(B) Inversely
(C) equal
(D) None of these
Answer
Correct option is B

107. PNN/GRNN networks have one neuron for each point in the training file, While RBF network have a variable number of neurons that is usually
(A) less than the number of training points.
(B) greater than the number of training points
(C) equal to the number of training points
(D) None of these
Answer
Correct option is A

108. Which network is more accurate when the size of training set between small to medium?
(A) PNN/GRNN
(B) RBF
(C) K-means clustering
(D) None of these
Answer
Correct option is A

109. What is/are true about RBF network?
(A) A kind of supervised learning
(B) Design of NN as curve fitting problem
(C) Use of multidimensional surface to interpolate the test data
(D) All of these
Answer
Correct option is D

110. Application of CBR
(A) Design
(B) Planning
(C) Diagnosis
(D) All of these
Answer
Correct option is A

111. What is/are advantages of CBR?
(A) A local approx. is found for each test case
(B) Knowledge is in a form understandable to human
(C) Fast to train
(D) All of these
Answer
Correct option is D

112 In k-NN algorithm, given a set of training examples and the value of k < size of training set (n), the algorithm predicts the class of a test example to be the. What is/are advantages of CBR?
(A) Least frequent class among the classes of k closest training examples.
(B) Most frequent class among the classes of k closest training examples.
(C) Class of the closest point.
(D) Most frequent class among the classes of the k farthest training examples.
Answer
Correct option is B

113. Which of the following statements is true about PCA?
(i) We must standardize the data before applying PCA.
(ii) We should select the principal components which explain the highest variance
(iii) We should select the principal components which explain the lowest variance
(iv) We can use PCA for visualizing the data in lower dimensions
(A) (i), (ii) and (iv).
(B) (ii) and (iv)
(C) (iii) and (iv)
(D) (i) and (iii)
Answer
Correct option is A

114. Genetic algorithm is a
(A) Search technique used in computing to find true or approximate solution to optimization and search problem
(B) Sorting technique used in computing to find true or approximate solution to optimization and sort problem
(C) Both A & B
(D) None of these
Answer
Correct option is A

115. GA techniques are inspired by _____ biology.
(A) Evolutionary
(B) Cytology
(C) Anatomy
(D) Ecology
Answer
Correct option is A

116. When would the genetic algorithm terminate?
(A) Maximum number of generations has been produced
(B) Satisfactory fitness level has been reached for the population.
(C) Both A & B
(D) None of these
Answer
Correct option is C

117. The algorithm operates by iteratively updating a pool of hypotheses, called the
(A) Population
(B) Fitness
(C) None of these
Answer
Correct option is A

118. What is the correct representation of GA?
(A) GA(Fitness, Fitness_threshold, p)
(B) GA(Fitness, Fitness_threshold, p, r)
(C) GA(Fitness, Fitness_threshold, p, r, m)
(D) GA(Fitness, Fitness_threshold)
Answer
Correct option is C

119. Genetic operators includes
(A) Crossover
(B) Mutation
(C) Both A & B
(D) None of these
Answer
Correct option is C

120. Produces two new offspring from two parent string by copying selected bits from each parent is called
(A) Mutation
(B) Inheritance
(C) Crossover
(D) None of these
Answer
Correct option is C

121. Each schema the set of bit strings containing the indicated as
(A) 0s, 1s
(B) only 0s
(C) only 1s
(D) 0s, 1s, *s
Answer
Correct option is D

122. 0*10 represents the set of bit strings that includes exactly
(A) 0010, 0110
(B) 0010, 0010
(C) 0100, 0110
(D) 0100, 0010
Answer
Correct option is A

123. Correct (h) is the percent of all training examples correctly classified by hypothesis h. then Fitness function is equal to
(A) Fitness (h) = (correct (h)) ^{2}
(B) Fitness (h) = (correct (h)) ^{3}
(C) Fitness (h) = (correct (h))
(D) Fitness (h) = (correct (h)) ^{4}
Answer
Correct option is A

124. Statement: Genetic Programming individuals in the evolving population are computer programs rather than bit strings.
(A) True
(B) False
Answer
Correct option is A

125. _____ evolution over many generations was directly influenced by the experiences of individual organisms during their lifetime
(A) Baldwin
(B) Lamarckian
(C) Bayes
(D) None of these
Answer
Correct option is B

126. Search through the hypothesis space cannot be characterized. Why?
(A) Hypotheses are created by crossover and mutation operators that allow radical changes between successive generations
(B) Hypotheses are not created by crossover and mutation operators.
(D) None of these
Answer
Correct option is A

127. ILP stand for
(A) Inductive Logical programming
(B) Inductive Logic Programming
(C) Inductive Logical Program
(D) Inductive Logic Program
Answer
Correct option is B

128. What is/are the requirement for the Learn-One-Rule method?
(A) Input, accepts a set of +ve and -ve training examples.
(B) Output, delivers a single rule that covers many +ve examples and few -ve.
(C) Output rule has a high accuracy but not necessarily a high coverage.
(D) A & B
(E) A, B & C
Answer
Correct option is E

129. _____ is any predicate (or its negation) applied to any set of terms.
(A) Literal
(B) Null
(C) Clause
(D) None of these
Answer
Correct option is A

130. Ground literal is a literal that
(A) Contains only variables
(B) does not contains any functions
(C) does not contains any variables
(D) Contains only functions
Answer
Correct option is C

131. _____ emphasizes learning feedback that evaluates the learner's performance without providing standards of correctness in the form of behavioural targets.
(A) Reinforcement learning
(B) Supervised Learning
(C) None of these
Answer
Correct option is A

132. Features of Reinforcement learning
(A) Set of problem rather than set of techniques
(B) RL is training by reward and punishments.
(C) RL is learning from trial and error with the world.
(D) All of these
Answer
Correct option is D

133. Which type of feedback used by RL?
(A) Purely Instructive feedback
(B) Purely Evaluative feedback
(C) Both A & B
(D) None of these
Answer
Correct option is B

134. What is/are the problem solving methods for RL?

(A) Dynamic programming
(B) Monte Carlo Methods
(C) Temporal-difference learning
(D) All of these
Answer
Correct option is D

135. The FIND-S Algorithm
(A) Starts with starts from the most specific hypothesis
(B) It considers negative examples only.
(C) It considers both negative and positive examples.
(D) None of these
Correct option is A

136. The hypothesis space has a general-to-specific ordering of hypotheses, and the search can be efficiently organized by taking advantage of a naturally occurring structure over the hypothesis space
(A) TRUE
(B) FALSE
Answer
Correct option is A

137. The Version space is:
(A) The subset of all hypotheses is called the version space with respect to the hypothesis space H and the training examples D, because it contains all plausible versions of the target concept.
(B) The version space consists of only specific hypotheses.
(C) None of these
Answer
Correct option is A

138. The Candidate-Elimination Algorithm
(A) The key idea in the Candidate-Elimination algorithm is to output a description of the set of all hypotheses consistent with the training examples.
(B) Candidate-Elimination algorithm computes the description of this set without explicitly enumerating all of its members.
(C) This is accomplished by using the more-general-than partial ordering and maintaining a compact representation of the set of consistent hypotheses.
(D) All of these
Answer
Correct option is D

139. Concept learning is basically acquiring the definition of a general category from given sample positive and negative training examples of the category.
(A) TRUE
(B) FALSE

Answer
Correct option is A

140. The hypothesis h1 is more-general-than hypothesis h2 (h1 > h2) if and only if h1≥h2 is true and h2≥h1 is false. We also say h2 is more-specific-than h1
(A) The statement is true
(B) The statement is false
(C) We cannot conclude.
(D) None of these
Answer
Correct option is A

141. The List-Then-Eliminate Algorithm
(A) The List-Then-Eliminate algorithm initializes the version space to contain all hypotheses in H, then eliminates any hypothesis found inconsistent with any training example.
(B) The List-Then-Eliminate algorithm not initializes to the version space.
(C) None of these
Answer
Correct option is A

142. What will take place as the agent observes its interactions with the world?
(A) Learning
(B) Hearing
(C) Perceiving
(D) Speech
Answer
Correct option is A

143. Which modifies the performance element so that it makes better decision?Performance element
(A) Performance element
(B) Changing element
(C) Learning element
(D) None of the mentioned
Answer
Correct option is C

144. Any hypothesis found to approximate the target function well over a sufficiently large set of training examples will also approximate the target function well over other unobserved example is called:
(A) Inductive Learning Hypothesis
(B) Null Hypothesis
(C) Actual Hypothesis
(D) None of these
Answer
Correct option is A

145. Feature of ANN in which ANN creates its own organization or representation of information it receives during learning time is
(A) Adaptive Learning
(B) Self Organization
(C) What-If Analysis
(D) Supervised Learning

Answer
Correct option is B

146. How the decision tree reaches its decision?
(A) Single test
(B) Two test
(C) Sequence of test
(D) No test

Answer
Correct option is C

147. Which of the following is a disadvantage of decision trees?
(A) Factor analysis
(B) Decision trees are robust to outliers
(C) Decision trees are prone to be overfit
(D) None of the above

Answer
Correct option is C

148. Tree/Rule based classification algorithms generate which rule to perform the classification.
(A) if-then.
(B) while.
(C) do while.
(D) switch.

Answer
Correct option is A

149. What is Gini Index?
(A) It is a type of index structure
(B) It is a measure of purity
(C) None of the options

Answer
Correct option is A

150. What is not a RNN in machine learning?
(A) One output to many inputs
(B) Many inputs to a single output
(C) RNNs for nonsequential input

(D) Many inputs to many outputs

Answer

Correct option is A

151. Which of the following sentences are correct in reference to Information gain?

(A) It is biased towards multi-valued attributes

(B) ID3 makes use of information gain

(C) The approach used by ID3 is greedy

(D) All of these

Answer

Correct option is D

152. A Neural Network can answer

(A) For Loop questions

(B) what-if questions

(C) IF-The-Else Analysis Questions

(D) None of these

Answer

Correct option is B

153. Artificial neural network used for

(A) Pattern Recognition

(B) Classification

(C) Clustering

(D) All

Answer

Correct option is D

154. Which of the following are the advantage/s of Decision Trees?

(E) Possible Scenarios can be added

(F) Use a white box model, If given result is provided by a model

(G) Worst, best and expected values can be determined for different scenarios

(H) All of the mentioned

Answer

Correct option is D

155. What is the mathematical likelihood that something will occur?

(A) Classification

(B) Probability

(C) Naïve Bayes Classifier

(D) None of the other answers are correct.

Answer

Correct option is C

156. What does the Bayesian network provides?

(A) Complete description of the domain
(B) Partial description of the domain
(C) Complete description of the problem
(D) None of the mentioned

Answer
Correct option is C

157. Where does the Bayes rule can be used?
(A) Solving queries
(B) Increasing complexity
(C) Decreasing complexity
(D) Answering probabilistic query

Answer
Correct option is D

158. How many terms are required for building a Bayes model?

(A) 2
(B) 3
(C) 4
(D) 1

Answer
Correct option is B

159. What is needed to make probabilistic systems feasible in the world?
(A) Reliability
(B) Crucial robustness
(C) Feasibility
(D) None of the mentioned

Answer
Correct option is B

160. It was shown that the Naive Bayesian method
(A) Can be much more accurate than the optimal Bayesian method
(B) Is always worse off than the optimal Bayesian method
(C) Can be almost optimal only when attributes are independent
(D) Can be almost optimal when some attributes are dependent

Answer
Correct option is C

161. What is the consequence between a node and its predecessors while creating Bayesian network?
(A) Functionally dependent
(B) Dependant
(C) Conditionally independent
(D) Both Conditionally dependant & Dependant

Answer
Correct option is C

162. How the compactness of the Bayesian network can be described?
(A) Locally structured
(B) Fully structured
(C) Partial structure
(D) All of the mentioned
Answer
Correct option is A

163. How the entries in the full joint probability distribution can be calculated?
(A) Using variables
(B) Using information
(C) Both Using variables & information
(D) None of the mentioned
Answer
Correct option is B

164. How the Bayesian network can be used to answer any query?
(A) Full distribution
(B) Joint distribution
(C) Partial distribution
(D) All of the mentioned
Answer
Correct option is B

165. Sample Complexity is
(A) The sample complexity is the number of training-samples that we need to supply to the algorithm, so that the function returned by the algorithm is within an arbitrarily small error of the best possible function, with probability arbitrarily close to 1
(B) How many training examples are needed for learner to converge to a successful hypothesis.
(C) All of these
Answer
Correct option is C

166. PAC stands for
(A) Probability Approximately Correct
(B) Probability Applied Correctly
(C) Partition Approximately Correct
Answer
Correct option is A

167. Which of the following will be true about k in k-NN in terms of variance
(A) When you increase the k the variance will increases

(B) When you decrease the k the variance will increases
(C) Can't say
(D) None of these

Answer
Correct option is B

168. Which of the following option is true about k-NN algorithm?
(A) It can be used for classification
(B) It can be used for regression
(C) It can be used in both classification and regression

Answer
Correct option is C

169. In k-NN it is very likely to overfit due to the curse of dimensionality. Which of the following option would you consider to handle such problem? 1. Dimensionality Reduction, 2. Feature selection
(A) 1
(B) 2
(C) 1 and 2
(D) None of these

Answer
Correct option is C

170. When you find noise in data which of the following option would you consider in k-NN
(A) I will increase the value of k
(B) I will decrease the value of k
(C) Noise can not be dependent on value of k
(D) None of these
Answer
Correct option is A

171. Which of the following will be true about k in k-NN in terms of Bias?
(A) When you increase the k the bias will be increases
(B) When you decrease the k the bias will be increases
(C) Can't say
(D) None of these

Answer
Correct option is A

172. What is used to mitigate overfitting in a test set?
(A) Overfitting set
(B) Training set
(C) Validation dataset
(D) Evaluation set

Answer
Correct option is C

173. A radial basis function is a
(A) Activation function
(B) Weight
(C) Learning rate
(D) none

Answer
Correct option is A

174. Mistake Bound is

(D) How many training examples are needed for learner to converge to a successful hypothesis.
(E) How much computational effort is needed for a learner to converge to a successful hypothesis
(F) How many training examples will the learner misclassify before conversing to a successful hypothesis
(G) None of these

Answer
Correct option is C

175. All of the following are suitable problems for genetic algorithms EXCEPT
(A) dynamic process control
(B) pattern recognition with complex patterns
(C) simulation of biological models
(D) simple optimization with few variables

Answer
Correct option is D

176. Adding more basis functions in a linear model... (Pick the most probably option)
(A) Decreases model bias
(B) Decreases estimation bias
(C) Decreases variance
(D) Doesn't affect bias and variance

Answer
Correct option is A

177. Which of these are types of crossover
(A) Single point
(B) Two point
(C) Uniform
(D) All of these

Answer
Correct option is D

178. A feature F1 can take certain value: A, B, C, D, E, & F and represents grade of students from a college. Which of the following statement is true in following case?

(A) Feature F1 is an example of nominal variable.

(B) Feature F1 is an example of ordinal variable.

(C) It doesn't belong to any of the above category.

Answer

Correct option is B

179. You observe the following while fitting a linear regression to the data: As you increase the amount of training data, the test error decreases and the training error increases. The train error is quite low (almost what you expect it to), while the test error is much higher than the train error. What do you think is the main reason behind this behaviour? Choose the most probable option.

(A) High variance

(B) High model bias

(C) High estimation bias

(D) None of the above

Answer

Correct option is C

180. Genetic algorithms are heuristic methods that do not guarantee an optimal solution to a problem

(A) TRUE

(B) FALSE

Answer

Correct option is A

181. Which of the following statements about regularization is not correct?

(A) Using too large a value of lambda can cause your hypothesis to underfit the data.

(B) Using too large a value of lambda can cause your hypothesis to overfit the data.

(C) Using a very large value of lambda cannot hurt the performance of your hypothesis.

(D) None of the above

Answer

Correct option is A

182. Consider the following: (a) Evolution (b) Selection (c) Reproduction (d) Mutation Which of the following are found in genetic algorithms?

(A) All

(B) a, b, c

(C) a, b

(D) b, d

Answer

Correct option is A

183. Genetic Algorithm are a part of
(A) Evolutionary Computing
(B) inspired by Darwin's theory about evolution - "survival of the fittest"
(C) are adaptive heuristic search algorithm based on the evolutionary ideas of natural selection and genetics
(D) All of the above

Answer
Correct option is D

184. Genetic algorithms belong to the family of methods in the
(A) artificial intelligence area
(B) optimization area.
(C) complete enumeration family of methods
(D) Non-computer based (human) solutions area

Answer
Correct option is A

185. For a two player chess game, the environment encompasses the opponent
(A) True
(B) False

Answer
Correct option is A

186. Which among the following is not a necessary feature of a reinforcement learning solution to a learning problem?
(A) exploration versus exploitation dilemma
(B) trial and error approach to learning
(C) learning based on rewards
(D) representation of the problem as a Markov Decision Process

Answer
Correct option is D

187. Which of the following sentence is FALSE regarding reinforcement learning
(A) It relates inputs to outputs.
(B) It is used for prediction.
(C) It may be used for interpretation.
(D) It discovers causal relationships.

Answer
Correct option is D

188. The EM algorithm is guaranteed to never decrease the value of its objective function on any iteration
(A) TRUE
(B) FALSE

Answer

Correct option is A

189. Consider the following modification to the tic-tac-toe game: at the end of game, a coin is tossed and the agent wins if a head appears regardless of whatever has happened in the game.Can reinforcement learning be used to learn an optimal policy of playing Tic-Tac-Toe in this case?

(A) Yes
(B) No

Answer
Correct option is B

190. Out of the two repeated steps in EM algorithm, the step 2 is _____

(A) the maximization step
(B) the minimization step
(C) the optimization step
(D) the normalization step

Answer
Correct option is A

191. Suppose the reinforcement learning player was greedy, that is, it always played the move that brought it to the position that it rated the best. Might it learn to play better, or worse, than a non greedy player?

(A) Worse
(B) Better

Answer
Correct option is B

192. A chess agent trained by using Reinforcement Learning can be trained by playing against a copy of the same agent.

(A) True
(B) False

Answer
Correct option is A

193. The EM iteration alternates between performing an expectation (E) step, which creates a function for the expectation of the log-likelihood evaluated using the current estimate for the parameters, and a maximization (M) step, which computes parameters maximizing the expected log-likelihood found on the E step.

(A) TRUE
(B) FALSE

Answer
Correct option is A

194. Expectation–maximization (EM) algorithm is an

(A) Iterative

(B) Incremental
(C) None
Answer
Correct option is A

195. Feature need to be identified by using Well Posed Learning Problem:
(A) Class of tasks
(B) Performance measure
(C) Training experience
(D) All of these
Answer
Correct option is D

196. A computer program that learns to play checkers might improve its performance as:
(A) Measured by its ability to win at the class of tasks involving playing checkers
(B) Experience obtained by playing games against itself.
(C) Both a & b
(D) None of these
Answer
Correct option is C
197. Learning symbolic representations of concepts known as:
(A) Artificial Intelligence
(B) Machine Learning
(C) Both a & b
(D) None of these
Answer
Correct option is A
198. The field of study that gives computers the capability to learn without being explicitly programmed _____
(A) Machine Learning
(B) Artificial Intelligence
(C) Deep Learning
(D) Both a & b
Answer
Correct option is A

199. The autonomous acquisition of knowledge through the use of computer programs is called _____
(A) Artificial Intelligence
(B) Machine Learning
(C) Deep learning
(D) All of these
Answer
Correct option is B

200. Learning that enables massive quantities of data is known as
(A) Artificial Intelligence
(B) Machine Learning
(C) Deep learning
(D) All of these
Answer
Correct option is B
201. A different learning method does not include
(A) Memorization
(B) Analogy
(C) Deduction
(D) Introduction
Answer
Correct option is D
202. Types of learning used in machine learning.
(A) Supervised
(B) Unsupervised
(C) Reinforcement
(D) All of these
Answer
Correct option is D
203. A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience
(A) Supervised learning problem
(B) Un Supervised learning problem
(C) Well posed learning problem
(D) All of these
Answer
Correct option is C
204. Which of the following is a widely used and effective machine learning algorithm based on the idea of bagging?
(A) Decision Tree
(B) Regression
(C) Classification
(D) Random Forest
Answer
Correct option is D
205. How many types are available in machine learning?
(A) 1
(B) 2
(C) 3
(D) 4
Answer

Correct option is C

205. A model can learn based on the rewards it received for its previous action is known as:

(A) Supervised learning
(B) Unsupervised learning
(C) Reinforcement learning
(D) Concept learning

Answer

Correct option is C

206. A subset of machine learning that involves systems that think and learn like humans using artificial neural networks.

(A) Artificial Intelligence
(B) Machine Learning
(C) Deep Learning
(D) All of these

Answer

Correct option is C

207. A learning method in which a training data contains a small amount of labeled data and a large amount of unlabeled data is known as _____

(A) Supervised Learning
(B) Semi Supervised Learning
(C) Unsupervised Learning
(D) Reinforcement Learning

Answer

Correct option is C

208. Methods used for the calibration in Supervised Learning

(A) Platt Calibration
(B) Isotonic Regression
(C) All of these
(D) None of above

Answer

Correct option is C

209. The basic design issues for designing a learning systems.

(A) Choosing the Training Experience
(B) Choosing the Target Function
(C) Choosing a Function Approximation Algorithm
(D) Estimating Training Values
(E) All of these

Answer

Correct option is E

210. In Machine learning the module that must solve the given performance task is known as:

(A) Critic
(B) Generalizer
(C) Performance system

(D) All of these

Answer

Correct option is C

211. A learning method that is used to solve a particular computational program, multiple models such as classifiers or experts are strategically generated and combined is called as _____

(A) Supervised Learning
(B) Semi Supervised Learning
(C) Unsupervised Learning
(D) Reinforcement Learning
(E) Ensemble learning

Answer

Correct option is E

212. In a learning system the component that takes as takes input the current hypothesis (currently learned function) and outputs a new problem for the Performance System to explore.

(A) Critic
(B) Generalizer
(C) Performance system
(D) Experiment generator
(E) All of these

Answer

Correct option is D

213. Learning method that is used to improve the classification, prediction, function approximation etc of a model.

(A) Supervised Learning
(B) Semi Supervised Learning
(C) Unsupervised Learning
(D) Reinforcement Learning
(E) Ensemble learning

Answer

Correct option is E

214. In a learning system the component that takes as input the history or trace of the game and produces as output a set of training examples of the target function is known as:

(A) Critic
(B) Generalizer
(C) Performance system
(D) All of these

Answer

Correct option is A

215. The most common issue when using ML is_____

(A) Lack of skilled resources
(B) Inadequate Infrastructure
(C) Poor Data Quality

(D) None of these

Answer

Correct option is C

216. How to ensure that your model is not over fitting

(A) Cross validation

(B) Regularization

(C) All of these

(D) None of these

Answer

Correct option is C

217. A way to ensemble multiple classifications or regression model.

(A) Stacking

(B) Bagging

(C) Blending

(D) Boosting

Answer

Correct option is A

218. How well a model is going to generalize in new environment is known as_____

(A) Data Quality

(B) Transparent

(C) Implementation

(D) None of these

Answer

Correct option is B

219. Common classes of problems in machine learning is _____

(A) Classification

(B) Clustering

(C) Regression

(D) All of these

Answer

Correct option is D

220. Which of the following is a widely used and effective machine learning algorithm based on the idea of bagging?

(A) Decision Tree

(B) Regression

(C) Classification

(D) Random Forest

Answer

Correct option is D

221. Cost complexity pruning algorithm is used in?

(A) CART

(B) C4.5

(C) ID3

(D) All of these.

Answer
Correct option is A
222. Which one of these is not a tree based learner?
(A) CART
(B) C4.5
(C) ID3
(D) Bayesian Classifier
Answer
Correct option is D
223. Which one of these is a tree based learner?
(A) Rule based
(B) Bayesian Belief Network
(C) Bayesian classifier
(D) Random Forest
Answer
Correct option is D
224. What is the approach of basic algorithm for decision tree induction?
(A) Greedy
(B) Top Down
(C) Procedural
(D) Step by Step
Answer
Correct option is A
225. Which of the following classifications would best suit the student performance classification systems?
(A) If-.then-analysis
(B) Market-basket analysis
(C) Regression analysis
(D) Cluster analysis
Answer
Correct option is A
226. What are two steps of tree pruning work?
(A) Pessimistic pruning and Optimistic pruning
(B) Post pruning and Pre pruning
(C) Cost complexity pruning and time complexity pruning
(D) None of these.
Answer
Correct option is B
227. How will you counter over-fitting in decision tree?
(A) By pruning the longer rules
(B) By creating new rules
(C) Both By pruning the longer rules' and ' By creating new rules'
(D) None of these.
Answer

Correct option is A

228. Which of the following sentences are true?

(A) In pre-pruning a tree is 'pruned' by halting its construction early

(B) A pruning set of class labeled tuples is used to estimate cost complexity.

(C) The best pruned tree is the one that minimizes the number of encoding

(D) bits.

(E) All of the above.

Answer

Correct option is D

229. Which of the following is a disadvantage of decision trees?

(A) Factor analysis

(B) Decision trees are robust to outliers

(C) Decision trees are prone to be over fit

(D) None of the above

Answer

Correct option is C

230. In which of the following scenario a gain ratio is preferred over Information Gain?

(A) When a categorical variable has very large number of category

(B) When a categorical variable has very small number of category

(C) Number of categories is the not the reason

(D) None of these

Answer

Correct option is A

231. Major pruning techniques used in decision tree are

(A) Minimum error

(B) Smallest tree

(C) Both a & b

(D) None of these

Answer

Correct option is B

232. What does the central limit theorem state?

(A) If the sample size increases sampling distribution must approach normal distribution

(B) If the sample size decreases then the sample distribution must approach normal distribution.

(C) If the sample size increases then the sampling distributions much approach an exponential distribution.

(D) If the sample size decreases then the sampling distributions much approach an exponential distribution.

Answer

Correct option is A

233. The difference between the sample value expected and the estimates value of the parameter is called as?

(A) Bias

(B) Error

(C) Contradiction
(D) Difference

Answer
Correct option is A

234. In which of the following types of sampling the information is carried out under the opinion of an expert?

(A) Quota sampling
(B) Convenience sampling
(C) Purposive sampling
(D) Judgment sampling

Answer
Correct option is D

235. Which of the following is a subset of population?

(A) Distribution
(B) Sample
(C) Data
(D) Set

Answer
Correct option is B

236. The sampling error is defined as?

(A) Difference between population and parameter
(B) Difference between sample and parameter
(C) Difference between population and sample
(D) Difference between parameter and sample

Answer
Correct option is C

237. Machine learning is interested in the best hypothesis h from some space H, given observed training data D. Here best hypothesis means

(A) Most general hypothesis
(B) Most probable hypothesis
(C) Most specific hypothesis
(D) None of these

Answer
Correct option is B

238. Practical difficulties with Bayesian Learning :

(A) Initial knowledge of many probabilities is required
(B) No consistent hypothesis
(C) Hypotheses make probabilistic predictions
(D) None of these

Answer
Correct option is A

239. Bayes' theorem states that the relationship between the probability of the hypothesis before getting the evidence P(H) and the probability of the hypothesis after getting the evidence P(H|E) is

(A) [P(E|H)P(H)] / P(E)
(B) [P(E|H) P(E)] / P(H)
(C) [P(E) P(H)] / P(E|H)
(D) None of these

Answer
Correct option is A

240. A doctor knows that Cold causes fever 50% of the time. Prior probability of any patient having cold is 1/50,000. Prior probability of any patient having fever is 1/20
If a patient has fever, what is the probability he/she has cold?

(A) P(C/F)= 0.0003
(B) P(C/F)=0.0004
(C) P(C/F)= 0.0002
(D) P(C/F)=0.0045

Answer
Correct option is C

241. Which of the following will be true about k in K-Nearest Neighbor in terms of Bias?

(A) When you increase the k the bias will be increases
(B) When you decrease the k the bias will be increases
(C) Can't say
(D) None of these

Answer
Correct option is A

242. When you find noise in data which of the following option would you consider in K-Nearest Neighbor?

(A) I will increase the value of k
(B) I will decrease the value of k
(C) Noise cannot be dependent on value of k
(D) None of these

Answer
Correct option is A

243. In K-Nearest Neighbor it is very likely to overfit due to the curse of dimensionality. Which of the following option would you consider to handle such problem?
1) Dimensionality Reduction
2) Feature selection

(A) 1
(B) 2
(C) 1 and 2
(D) None of these

Answer
Correct option is C

244. Radial basis functions is closely related to distance-weighted regression, but it is

(A) lazy learning
(B) eager learning
(C) concept learning
(D) none of these

Answer
Correct option is B

245. Radial basis function networks provide a global approximation to the target function, represented by _____ of many local kernel functions.

(A) a series combination
(B) a linear combination
(C) a parallel combination
(D) a non linear combination

Answer
Correct option is B

246. The most significant phase in a genetic algorithm is

(A) Crossover
(B) Mutation
(C) Selection
(D) Fitness function

Answer
Correct option is A

247. The crossover operator produces two new offspring from

(A) Two parent strings, by copying selected bits from each parent
(B) One parent strings, by copying selected bits from selected parent
(C) Two parent strings, by copying selected bits from one parent
(D) None of these

Answer
Correct option is A

248. Mathematically characterize the evolution over time of the population within a GA based on the concept of

(A) Schema
(B) Crossover
(C) Don't care
(D) Fitness function

Answer
Correct option is A

249. In genetic algorithm process of selecting parents which mate and recombine to create off-springs for the next generation is known as:

(A) Tournament selection
(B) Rank selection
(C) Fitness sharing
(D) Parent selection

Answer

Correct option is D

250. Crossover operations are performed in genetic programming by replacing

(A) Randomly chosen sub tree of one parent program by a sub tree from the other parent program.

(B) Randomly chosen root node tree of one parent program by a sub tree from the other parent program

(C) Randomly chosen root node tree of one parent program by a root node tree from the other parent program

(D) None of these

Answer

Correct option is A

**********Best of Luck**********

AKTU EXAM 19-20

Machine Learning Solved MCQ

Answer Key

Question	Answer	Question	Answer	Question	Answer
1	A	**26**	C	**51**	A
2	D	**27**	C	**52**	A
3	C	**28**	B	**53**	A
4	D	**29**	B	**54**	C
5	D	**30**	C	**55**	D
6	B or C	**31**	C	**56**	A
7	D	**32**	D	**57**	C
8	A	**33**	C	**58**	B
9	A	**34**	A	**59**	D
10	A	**35**	C	**60**	A
11	A	**36**	A	**61**	C
12	A	**37**	D	**62**	D
13	C	**38**	C	**63**	C
14	B	**39**	D	**64**	D
15	B	**40**	B	**65**	D
16	C	**41**	A	**66**	C
17	B	**42**	D	**67**	A
18	B	**43**	B	**68**	A
19	C	**44**	C	**69**	B
20	A	**45**	B	**70**	A
21	B	**46**	C		
22	B	**47**	B		
23	A	**48**	D		
24	A	**49**	D		
25	B	**50**	D		

AKTU EXAM 19-20
Machine Learning Solved MCQ
Highlighted Option is Correct Answer

Note: Attempt all questions. The question paper contains 70 MCQ type questions. Each question carries equal marks. Select the answer and fill the bubble corresponding to that question in the attached OMR sheet.

1. What is Machine learning?
   (A) The autonomous acquisition of knowledge through the use of computer programs
   (B) The autonomous acquisition of knowledge through the use of manual programs
   (C) The selective acquisition of knowledge through the use of computer programs
   (D) The selective acquisition of knowledge through the use of manual programs

2. Which of the factors affect the performance of learner system does not include?
   (A) Representation scheme used
   (B) Training scenario
   (C) Type of feedback
   (D) Good data structures

3. Which of the following statements is/are true about "Type-1" and "Type-2" errors?
   (i) Type1 is known as false positive and Type2 is known as false negative.
   (ii) Type1 is known as false negative and Type2 is known as false positive.
   (iii) Type1 error occurs when we reject a null hypothesis when it is true.
   (A) Only i
   (B) Only iii
   (C) i and iii
   (D) ii and iii

4. How do you handle missing or corrupted data in a dataset?
   (A) Drop missing rows or columns
   (B) Replace missing values with mean/median/mode
   (C) Assign a unique category to missing values
   (D) All of the above

5. Which is of the following option is true about FIND-S Algorithm
   (A) FIND-S Algorithm starts from the most specific hypothesis and generalize it by considering only positive examples.
   (B) FIND-S algorithm ignores negative examples.
   (C) FIND-S algorithm finds the most specific hypothesis within H that is consistent with the positive training examples.

(D) All of the above

6. Regarding bias and variance, which of the following statements are true? (Here 'high' and 'low' are relative to the ideal model.)
   (A) Models which overfit have a high bias.
   (B) Models which overfit have a low bias.
   (C) Models which underfit have a high variance.
   (D) None of these
7. Which of the following sentence is FALSE regarding regression?
   (A) It relates inputs to outputs.
   (B) It is used for prediction.
   (C) It may be used for interpretation.
   (D) It discovers causal relationships.
8. You observe the following while fitting a linear regression to the data: As you increase the amount of training data, the test error decreases, and the training error increases. The train error is quite low (almost what you expect it to), while the test error is much higher than the train error. What do you think is the main reason behind this behavior? Choose the most probable option.
   (A) High variance
   (B) High model bias
   (C) High estimation bias
   (D) None of the above
9. Adding more basis functions in a linear model... (pick the most probably option)
   (A) Decreases model bias
   (B) Decreases estimation bias
   (C) Decreases variance
   (D) Doesn't affect bias and variance
10. Which of the following will be true about k in k-NN in terms of Bias?
   (A) When you increase the k the bias will be increases
   (B) When you decrease the k the bias will be increases
   (C) Can't say
   (D) None of these
11. Which of the following distance measure do we use in case of categorical variables in k-NN?
   Hamming Distance
   Euclidean Distance
   Manhattan Distance
   (A) 1
   (B) 2
   (C) 3
   (D) 1,2 and 3
12. Imagine, you are working with "Analytics Vidhya" and you want to develop a machine learning algorithm which predicts the number of views on the articles.

Your analysis is based on features like author name, number of articles written by the same author on Analytics Vidhya in past and a few other features. Which of the following evaluation metric would you choose in that case?

Mean Square Error

Accuracy

F1 Score

(A) Only 1

(B) Only 2

(C) Only 3

(D) 1 and 3

13. At a certain university, 4% of men are over 6 feet tall and 1% of women are over 6 feet tall. The total student population is divided in the ratio 3:2 in favour of women. If a student is selected at random from among all those over six feet tall, what is the probability that the student is a woman?

(A) 2/5

(B) 3/5

(C) 3/11

(D) 1/100

14. Macromutation operator is also known as

(A) Headed Chicken

(B) Headless chicken

(C) SPX operator

(D) BLX operator

15. Choose the False Statement.

Gradient of a continuous and differentiable function

(A) is zero at a minimum

(B) is non-zero at a maximum

(C) is zero at a saddle point

(D) decreases as you get closer to the minimum

16. Computational complexity of Gradient descent is,

(A) linear in D

(B) linear in N

(C) polynomial in D

(D) dependent on the number of iterations

17. Let's say, you are using activation function X in hidden layers of neural network. At a neuron for any given input, you get the output as "-0.0001". Which of the following activation function could X represent?

(A) ReLU

(B) tanh

(C) SIGMOID

(D) None of these

18. Which of the following hyper parameter(s), when increased may cause random forest to over fit the data?

Number of Trees

Depth of Tree

Learning Rate

(A) Only 1

(B) Only 2

(C) 1 and 2

(D) 2 and 3

19. Which of the following is a disadvantage of decision trees?

(A) Factor analysis

(B) Decision trees are robust to outliers

(C) Decision trees are prone to be overfit

(D) None of the above

20. To find the minimum or the maximum of a function, we set the gradient to zero because:

(A) The value of the gradient at extrema of a function is always zero

(B) Depends on the type of problem

(C) Both A and B

(D) None of the above

21. In Delta Rule for error minimization

(A) weights are adjusted w.r.to change in the output

(B) Weights are adjusted w.r.to difference between desired output and actual output

(C) weights are adjusted w.r.to difference between input and output

(D) none of the above

22. Back propagation is a learning technique that adjusts weights in the neural network by propagating weight changes.

(A) Forward from source to sink

(B) Backward from sink to source

(C) Forward from source to hidden nodes

(D) Backward from sink to hidden nodes

23. Which of the following neural networks uses supervised learning?

(A) Multilayer perceptron

(B) Self organizing feature map

(C) Hopfield network

Choose the correct answer:

(A) A only

(B) B only

(C) A and B only

(D) A and C only

24. Which of the following sentences is incorrect in reference to Information gain?

(A) It is biased towards single-valued attributes

(B) It is biased towards multi-valued attributes

(C) ID3 makes use of information gain

(D) The approach used by ID3 is greedy

25. What are two steps of tree pruning work?

(A) Pessimistic pruning and

Optimistic pruning

(B) Post-pruning and Pre-pruning

(C) Cost complexity pruning and time complexity pruning

(D) None of the options

26. Which one of these is not a tree-based learner?

(A) CART

(B) ID3

(C) Bayesian classifier

(D) Random Forest

27. What is tree-based classifiers?

(A) Classifiers which form a tree with each attribute at one level

(B) Classifiers which perform series of condition checking with one attribute at a time

(C) Both a and b

(D) None of the options

28. Decision Nodes are represented by _____

(A) Disks

(B) Squares

(C) Circles

(D) Triangles

29. Previous probabilities in Bayes Theorem that are changed with help of new available information are classified as _____

(A) independent probabilities

(B) posterior probabilities

(C) interior probabilities

(D) dependent probabilities

30. Which of the following is true about Naive Bayes?

(A) Assumes that all the features in a dataset are equally important

(B) Assumes that all the features in a dataset are independent

(C) Both A and B

(D) None of the above options

31. The method in which the previously calculated probabilities are revised with new probabilities is classified as

(A) updating theorem

(B) revised theorem

(C) Bayes theorem

(D) dependency theorem

32. Which of the following is a widely used and effective machine learning algorithm based on the idea of bagging?

(A) Decision Tree

(B) Regression

(C) Classification

(D) Random Forest

33. Which of the following is a good test dataset characteristic?

(A) Large enough to yield meaningful results

(B) Is representative of the dataset as a whole

(C) Both A and B

(D) None of the above

34. What is the arity in case of crossover

operator in GA?

(A) Number of parents used for the operator

(B) Number of offspring used for the operator

(C) Both a and b

(D) None

35. Which of the following statements about regularization is not correct?

(A) Using too large a value of lambda can cause your hypothesis to underfit the data.

(B) Using too large a value of lambda can cause your hypothesis to overfit the data.

(C) Using a very large value of lambda cannot hurt the performance of your hypothesis.

(D) None of the above

36. You are given reviews of movies marked as positive, negative, and neutral. Classifying reviews of a new movie is an example of

(A) Supervised Learning

(B) Unsupervised Learning

(C) Reinforcement Learning

(D) None of these

37. Regarding bias and variance, which of the following statements are true? (Here 'high' and 'low' are relative to the ideal model.)

(i) Models which overfit have a high bias.

(ii) Models which overfit have a low bias.

(iii) Models which underfit have a high variance.

(iv) Models which underfit have a low variance

(A) (i) and (ii)

(B) (ii) and (iv)

(C) (iii) and (iv)

(D) None of these

38. What is the purpose of restricting hypothesis space in machine learning?

(A) can be easier to search

(B) May avoid overfit since they are usually simpler (e.g. linear or low order decision surface)

(C) Both above

(D) None of the above

39. Suppose, you got a situation where you find that your linear regression model is under fitting the data. In such situation which of the following options would you consider?

(A) You will add more features

(B) You will start introducing higher degree features

(C) You will remove some features

(D) Both a and b.

40. Consider a simple linear regression model with one independent variable

(X). The output variable is Y. The equation is : Y=aX+b, where a is the slope and b is the intercept. If we change the input variable (X) by 1 unit, by how much output variable (Y) will change?

(A) 1 unit

(B) By slope

(C) By intercept,

(D) None

41. You have generated data from a 3-degree polynomial with some noise. What do you expect of the model that was trained on this data using a 5-degree polynomial as function class?

(A) Low bias, high variance

(B) High bias, low variance.

(C) Low bias, low variance.

(D) High bias, low variance.

42. Genetic Algorithm are a part of

(A) Evolutionary Computing

(B) inspired by Darwin's theory about evolution - "survival of the fittest"

(C) are adaptive heuristic search algorithm based on the evolutionary ideas of natural selection and genetics

(D) All of the above

43. What are the 2 types of learning

(A) Improvised and un-improvised

(B) supervised and unsupervised

(C) Layered and unlayered

(D) None of the above

44. Unsupervised learning is

(A) learning without computers

(B) problem based learning

(C) learning from environment

(D) learning from teachers

45. In supervised learning

(A) classes are not predefined

(B) classes are predefined

(C) classes are not required

(D) classification is not done

46. Mutating a strain is:

(A) Changing all the genes in the strain.

(B) Removing one gene in the strain.

(C) Randomly changing one gene in the strain.

(D) Removing the strain from the population.

47. Genetic Algorithms are considered pseudo-random because they:

(A) Search the solution space in a random fashion.

(B) Search the solution space using the previous generation as a starting point.

(C) Have no knowledge of what strains are contained in the next generation.

(D) Use random numbers.

48. The three gene operators we have discussed can be thought of as:

(A) Crossover: Receiving the best genes from both parents.

(B) Mutation: Changing one gene so that the child is almost like the parent.

(C) Mirror: Changing a string of genes in the child so it is like a cousinto the parent.

(D) A and B only

49. If a population contains only one strain, you can introduce new strains by:
Using the Crossover operator.

(A) Injecting random strains into the population.

(B) Using the Mutation operator.

(C) B only

(D) B and C only

50. The efficiency of a Genetic Algorithm (how quickly it arrives at the best solution) is dependent upon:

(A) The initial conditions.

(B) The size of the population.

(C) The types of operators employed.

(D) All of the above

51. Which of the following methods do we use, to find the best fit line for data in Linear Regression?

(A) Least Square Error

(B) Maximum Likelihood

(C) Logarithmic Loss

(D) Both A and B

52. Among the following, which one is not "hyperparameter"?

(A) learning rate α

(B) number of layers L in the neural network

(C) activation values a[l]

(D) size of the hidden layers n[l]

53. (i) The deeper layers of a neural network are typically computing more complex features of the input than the earlier layers.

(ii) The earlier layers of a neural network are typically computing more complex features of the input than the deeper layers.

Which of the following option is correct?

(A) (i) is correct and (ii) is incorrect

(B) (i) is incorrect while (ii) is correct

(C) both are correct

(D) both are incorrect

54. There are certain functions with the following properties:

(i) To compute the function using a shallow network circuit, you will need a large network (where we measure size by the number of logic gates in the network)

(ii) To compute it using a deep network circuit, you need only an exponentially smaller

network.

Which of the following option is correct?

(A) (i) is correct and (ii) is incorrect

(B) (i) is incorrect while (ii) is correct

(C) both are correct

(D) both are incorrect

55. Factor Analysis involves:

(A) dimensionality reduction technique

(B) finding correlation among variables

(C) capturing maximum variance in the data with minimum number of variables

(D) All the above

56. Which of the following is way to reduce the skewness of a variable?

(A) Taking log of the skewed variable

(B) Dividing each value of skewed variable by its standard deviation.

(C) Normalizing the skewed variable

(D) Standardizing the skewed variable.

57. what causes overfitting?

(A) Large number of features in the data

(B) Noise in the data

(C) Both A and B

(D) None of the above

58. Given an image of a person,

(i) predicting the height of that person

(ii) finding whether the person is in happy, angry or sad mood.

type of ML problem is

(A) (i) is classification while (ii) is regression problem

(B) (ii) is classification while (i) is regression problem

(C) both are classification problem

(D) both are regression problem

59. what does fitness function represent to describe optimization problem?

(A) Objective function

(B) Scaling function

(C) Chromosome decoding function

(D) All of the above

60. which of the following algorithms is called Lazy Learner?

(A) KNN

(B) SVM

(C) Naïve Bayes

(D) Decision Tree

61. What are the main driving operators of GA?

(A) Selection

(B) Crossover

(C) Both a and b

(D) None of these

62. which of the following is true about bagging and boosting?
(A) Both are ensemble learning techniques
(B) Both combine the output of weak learners to make consistent predictions
(C) Both can be used to solve classification as well as regression problems
(D) All of the above

63. what causes underfitting?
(A) Less number of features in the data
(B) Less number of observations in the data
(C) Both a and b
(D) None of the above

64. The performance of GA is influenced by
(A) Population size
(B) Crossover rate
(C) Mutation rate
(D) All of the above

65. which of the following are main components of evolutionary computation?
(A) Initial population
(B) Fitness function
(C) Crossover, mutation and selection
(D) All of the above

66. which of the following statement(s) is/are true?
(A) Genetic algorithm mimic process from natural selection
(B) Chromosomes play vital roles in GA
(C) Both a and b
(D) Chromosomes can't be encoded

67. characteristics of individual is represented by
(A) Chromosomes
(B) Gray Code
(C) Initial population
(D) None of the above

68. what is the main concept of Evolutionary computation?
(A) Survival of the fittest
(B) Survival of the weakest
(C) Phenotype
(D) None of these

69. selective pressure is also known as
(A) Takeover Time
(B) candidate solution
(C) Proportionate time
(D) None of the above

70. Which selection strategy is susceptible to a high selection pressure and low population diversity?
(A) Roulette-wheel selection
(B) Rank based selection
(C) Tournament selection
(D) All of the above

******