BRAINWARE UNIVERSITY

Term End Examination 2021 - 22
Programme – Master of Technology in Computer Science & Engineering
Course Name – Machine Learning
Course Code - PEC-MCS201A
(Semester II)

Time allotted : 1 Hrs.15 Min. **Full Marks : 60**

[The figure in the margin indicates full marks.]

Group-A

(Multiple Choice Type Question) 1 x 60=60

Choose the correct alternative from the following :

(1) If a supermarket has L items, the number of possible itemsets is:

a) 2L-1 b) 2^L-1

c) L/2 d) L-1

(2) If A, B are two sets of items, and A⊆ B. Which of the following statement is always true?:

a) support(*A*) ≤ support(*B*) b) support(A) ≥ support(B)

c) support(*A*) = support(*B*) d) support(*A*) ≠ support(*B*)

(3) Consider three itemsets V1 = {tomato, potato, onion}, V2 = {tomato, potato}, V3 = {tomato}. Which of the following statements are correct?

a) support(V1) > support(V2) b) support(V3) > support(V2)

c) support(V1) > support(V3) d) none of the above

(4) In the following data table, what is the confidence of the rule b→c?

Transaction ID	Itemsets
1	{a, b, d, e}
2	{b, c, d}
3	{a, b, d, e}
4	{a, c, d, e}
5	{b, c, d, e}
6	{b, d, c}

7	{c, d}
8	{a, b, c}
9	{a, d, e}
10	{b, c}

a) 2/7
b) 3/7
c) 4/7
d) 5/7

(5) In the following data table, if the support threshold is (greater than or equal to) 0.2, and confidence threshold is (greater than or equal to) 0.9, valid association rules are:

Transaction ID	Itemsets
1	{a, b, d, e}
2	{b, c, d}
3	{a, b, d, e}
4	{a, c, d, e}
5	{b, c, d, e}
6	{b, d, c}
7	{c, d}
8	{a, b, c}
9	{a, d, e}
10	{b, c}

a) {a,b,d} $\rightarrow$ e
b) {a, b}$\rightarrow${d, e}
c) a$\rightarrow${b, d, e}
d) b $\rightarrow${a, d, e}

(6) Given the following training set for classification problem into two classes "fraud" and "normal". There are two attributes A1 and A2 taking values 0 or 1. Splitting on which attribute in the root of a decision tree will lead to highest information gain?

A1	A2	Class
1	0	fraud
1	1	fraud
1	1	fraud
1	0	normal
1	1	fraud
0	0	normal
0	0	normal
0	0	normal
1	1	normal
1	0	normal

a) 0
b) –(4/10)xlog(4/10)-(6/10)xlog(6/10)
c) –log(4/10)-log(6/10)
d) 1

(7) Given the following training set for classification problem into two classes "fraud" and "normal". There are two attributes A1 and A2 taking values 0 or 1. The Bayes classifier classifies the instance (A1=1, A2=1) into class?

A1	A2	Class
1	0	fraud
1	1	fraud
1	1	fraud
1	0	normal
1	1	fraud
0	0	normal
0	0	normal
0	0	normal
1	1	normal
1	0	normal

a) fraud
b) normal
c) there will be a tie
d) not enough information to classify

(8) The constraint in the primal optimization problem solved to obtain the hard margin optimal separating hyperplane is:

a) $y_i(W^TX_i+b) \geq 1$ for all i
b) $y_i(W^TX_i+b) \leq 1$ for all i
c) $(W^TX_i+b) \geq 1$ for all i
d) $(W^TX_i+b) \leq 1$ for all i

(9) In a hard margin SVM, support vectors lie -

a) inside the margin
b) on the margin
c) outside the margin
d) can be either inside or outside the margin

(10) In a hard margin SVM $W^TX+b=0$, suppose X_j's are the support vectors and α_j's the corresponding Lagrange multipliers, then which of the following statements are correct:

a) $W = \Sigma\, \alpha_j y_j X_j$
b) $\Sigma\, \alpha_j y_j = 0$
c) Either A or B
d) Both A and B

(11) A training input x is used for a perceptron learning rule. The desired output is t and the actual output is o. If learning rate is η, the weight update performed by the learning rule is describedby?

a) $w_i \leftarrow w_i + \eta(t - o)$
b) $w_i \leftarrow w_i + \eta(t - o)\, x$
c) $w_i \leftarrow \eta(t - o)\, x$
d) $w_i \leftarrow w_i + (t - o)\, x$

(12) Suppose we have n training examples x_i, $i=1...n$, whose desired outputs are t_i, $i=1...n$. The output of a perceptron for these training examples x_i 's are o_i, $i=1...n$. The error function minimised by the gradient descend perceptron learning algorithm is:

a)
b)
c)
d)

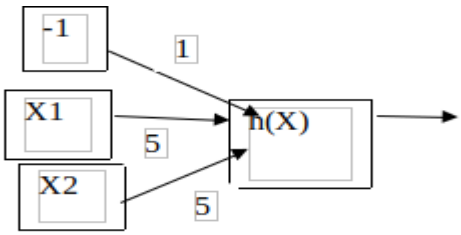
(13) The tanh activation function h(z)=21+e−2z−1 is:

a) Discontinuous and not differentiable
b) Discontinuous but differentiable
c) Continuous but not differentiable
d) Continuous and differentiable

(14) The neural network given bellow takes two binary valued inputs

x1,x2ϵ{0,1} and the activation function is the binary threshold function h(z)=1if *z*>0; 0 otherwise

Which of the following logical functions does it compute?

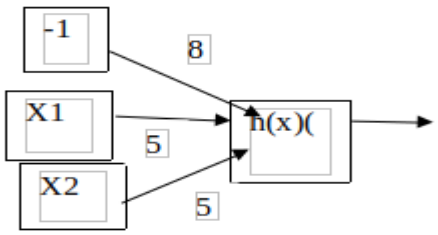


a) OR
b) AND
c) NAND
d) NOR

(15) The neural network given bellow takes two binary valued inputs

x1,x2ϵ{0,1}

and the activation function is the binary threshold function h(z)=1if z>0; 0 otherwise. Which of the following logical functions does it compute?



a) OR
b) AND
c) NAND
d) NOR

(16) Consider a set of five 2-dimensional points $p_1=(0, 0)$, $p_2=(0, 1)$, $p_3=(5, 8)$, $p_4=(5, 7)$, and $p_5=(0, 0.5)$. Euclidean distance is the distance function. Single linkage clustering is used to cluster the points into two clusters. The clusters are:

a) $\{p_1, p_2, p_3\}\{p_4, p_5\}$
b) $\{p_1, p_4, p_5\}\{p_2, p_3\}$
c) $\{p_1, p_2, p_5\}\{p_3, p_4\}$
d) $\{p_1, p_2, p_4\}\{p_3, p_5\}$

(17) Consider a set of five 2-dimensional points $p_1=(0, 0)$, $p_2=(0, 1)$, $p_3=(5, 8)$, $p_4=(5, 7)$, and $p_5=(0, 0.5)$. Euclidean distance is the distance function. The k-means algorithm is used to cluster the points into two clusters. The initial cluster centers are p_1 and p_4. The clusters after two iterations of k-means are:

a) $\{p_1, p_4, p_5\}\{p_2, p_3\}$
b) $\{p_1, p_2, p_5\}\{p_3, p_4\}$
c) $\{p_3, p_4, p_5\}\{p_1, p_2\}$
d) $\{p_1, p_2, p_4\}\{p_3, p_5\}$

(18) Considerx_1, x_2 to be the independent variables and y the dependent variable, which of the following represents a linear regression model?

a) $y = a_0 + a_1/x_1 + a_2/x_2$
b) $y = a_0 + a_1x_1 + a_2x_2$
c) $y = a_0 + a_1x_1 + a_2x_2^2$
d) $y = a_0 + a_1x_1^2 + a_2x_2$

(19) The linear regression model $y = a_0 + a_1x$is applied to the data in the table shown below. What is the value of the sum squared error function $S(a_0, a_1)$, when $a_0 = 1$, $a_1 = 2$?

x	y
0	1
0.5	1.9
1	2.5
1.25	3

a) 0.00
b) 0.25
c) 0.50
d) 0.51

(20) The linear regression model $y = a_0 + a_1x_1 + a_2x_2 + ... + a_px_p$ is to be fitted to a set of N training data points having p attributes each. Let X be $N \times (p+1)$ vectors of input values (augmented by 1's), Y be $N \times 1$ vector of target values, and θ be $(p+1) \times 1$ vector of parameter values ($a_0, a_1, a_2, \ldots, a_p$). If the sum squared error is minimized for obtaining the optimal regression model, which of the following equation holds?

a) $X^TX = Xy$
b) $X\theta = X^Ty$
c) $X^TX\theta = y$
d) $X^TX\theta = X^Ty$

(21) Which of the following is usually the last step in the data mining process?

a) Visualization
b) Preprocessing
c) Modeling
d) Deployment

(22) User rating given to a movie in a scale 1-10, can be considered as an attribute of type?

a) Nominal
b) Ordinal
c) Interval
d) Ratio

(23) Sales database of items in a supermarket can be considered as an example of:

a) Record data
b) Tree data
c) Graph data
d) None of the above

(24) Which of the following is an example of continuous attribute?

a) Weight of a person
b) Shoe size of a person
c) Gender of a person
d) None of the above

(25) If a record data matrix has reduced number of rows after a transformation, the transformation has performed:

a) Data sampling
b) Dimensionality reduction
c) Noise cleaning
d) Discretization

(26) An itemset satisfying the support criterion is known as:

a) Frequent itemset
b) Confident itemset
c) Accurate itemset
d) Reliable itemset

(27) Leaf nodes of a decision tree correspond to:

a) Attributes
b) Classes
c) Data instances
d) None of the above

(28) Which of the following criteria is used to decide which attribute to split next ina decision tree:

a) Support
b) Confidence
c) Entropy
d) Scatter

(29) The purpose of pruning a decision tree is:

a) improving training set classification accuracy
b) improving generalization performance
c) dimensionality reduction
d) tree balancing

(30) Maximum aposteriori classifier is also known as:

a) Decision tree classifier
b) Bayes classifier
c) Gaussian classifier
d) Maximum margin classifier

(31) Let A be an example, and C be a class. The probability P(C|A) is known as:

a) Apriori probability
b) Aposteriori probability
c) Class conditional probability
d) None of the above

(32) A bank classifies its customer into two classes "fraud" and "normal" based on their installment payment behavior. We know that the probability of a customer being being fraud is P(fraud) = 0.20, the probability of customer defaulting installment payment is P(default) = 0.40, and the probability that a fraud customer defaults in installment payment is P(default|fraud) = 0.80. What is the probability of a customer who defaults in payment being a fraud?

a) 0.8
b) 0.6
c) 0.4
d) 0.2

(33) Consider a binary classification problem with two classes C1 and C2. Class labels of ten other training set instances sorted in increasing order of their distance to an instance x is as follows: {C1, C2, C1, C2, C2, C2, C1, C2, C1, C2}. How will a K=7 nearest neighbor classifier classify x?

a) There will be a tie
b) C1
c) C2
d) Not enough information to classify

(34) In a hard margin support vector machine:

a) No training instances lie inside the margin
b) All the training instances lie inside the margin
c) Only few training instances lie inside the margin
d) None of the above

(35) The generalization constant C is used to tune the:

a) test error only
b) training error only
c) relative weightage to training and test error
d) none of the above

(36) A perceptron can correctly classify instances into two classes where the classes are:

a) Overlapping
b) Linearly separable
c) Non-linearly separable
d) None of the above

(37) Which of the following statement is true for a multilayered perceptron?

a) Output of all the nodes of a layer is input to all the nodes of the next layer
b) Output of all the nodes of a layer is input to all the nodes of the same layer
c) Output of all the nodes of a layer is input to all the nodes of the previous layer
d) Output of all the nodes of a layer is input to all the nodes of the output layer

(38) Overfitting is expected when we observe that?

a) With training iterations error on training set as well as test set decreases
b) With training iterations error on training set decreases but test set increases
c) With training iterations error on training set as well as test set increases
d) With training iterations training set as well as test set error remains constant

(39) Which of the following is an exploratory data mining technique?

a) Classification
b) Clustering
c) Regression
d) None of the above

(40) Which of the following clustering algorithm uses a dendogram?

a) Complete linkage clustering
b) K-means clustering
c) DBSCAN
d) None of the above

(41) Distance between two clusters in single linkage clustering is defined as:

a) Distance between the closest pair of points between the clusters
b) Distance between the furthest pair of points between the clusters
c) Distance between the most centrally located pair of points in the clusters
d) None of the above

(42) Which of the following is not true about K-means clustering algorithm?

a) It is a partitional clustering algorithm
b) The final cluster obtained depends on the choice of initial cluster centres
c) Number of clusters need to be specified in advance
d) It can generate non-convex cluster shapes

(43) Regression is used in:

a) predictive data mining
b) exploratory data mining
c) descriptive data mining
d) explanative data mining

(44) Regression finds out the model parameters which produces the least square error between -

a) input value and output value
b) input value and target value
c) output value and target value
d) model parameters and output value

(45) A time series prediction problem is often solved using?

a) Multivariate regression
b) Autoregression
c) Logistic regression
d) Sinusoidal regression

(46) What is global stability?

a) when both synaptic & activation dynamics are simultaneously used & are in equilibrium
b) when only synaptic dynamics in equilibrium
c) when only synaptic dynamics in equilibrium
d) none of the mentioned

(47) Who proposed the shunting activation model?

a) Rosenblatt
b) hopfield
c) perkel
d) grossberg

(48) What is true regarding backpropagation rule?

a) it is also called generalized delta rule
b) error in output is propagated backwards only to determine weight updates
c) there is no feedback of signal at any stage
d) all of the mentioned

(49) What are general limitations of back propagation rule?

a) local minima problem
b) slow convergence
c) scaling
d) all of the mentioned

(50) Does backpropagaion learning is based on gradient descent along error surface?

a) Yes
b) no
c) cannot be said
d) it depends on gradient descent but not error surface

(51) Supervised learning may be used for?

a) temporal learning
b) structural learning
c) both temporal & structural learning
d) none of the mentioned

(52) 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

(53) Which of the following is an advantage of simulation?

a) It can incorporate significant real-life complexity
b) It always results in optimal solutions
c) Simulation software requires special skills.
d) It solves problems in one pass with no iterations

(54) In which stage of the simulation methodology do you determine the system's boundaries and environment?

a) Constructing the simulation model
b) Defining the problem
c) Testing and validating the mode
d) Designing the experiment

(55) What happens if chain-termination mutation is in the S gene?

a) Cell lysis is blocked
b) Growth of cells containing low levels of packaging proteins is not allowed
c) The lysis of cells is not carried artificially
d) Packaging is not carried out efficiently

(56) Who initiated the idea of Soft Computing

a) Charles Darwin
b) Lofti A Zadeh
c) Rechenberg
d) McCulloch

(57) Artificial intelligence is

a) It uses machine-learning techniques. Here program can learn From past experience and adapt themselves to new situations
b) Computational procedure that takes some value as input and produces some value as output.
c) Science of making machines performs tasks that would require intelligence when performed by humans
d) None of these

(58) Falsification is

a) Modular design of a software application that facilitates the integration of new modules
b) Showing a universal law or rule to be invalid by providing a counter example
c) A set of attributes in a database table that refers to data in another table
d) None of these

(59) Extendible architecture is

a) Modular design of a software application that facilitates the integration of new modules
b) Showing a universal law or rule to be invalid by providing a counter example
c) A set of attributes in a database table that refers to data in another table
d) None of these

(60) Search space

a) The large set of candidate solutions possible for a problem
b) The information stored in a database that can be, retrieved with a single query.
c) Worth of the output of a machine learning program that makes it understandable for humans
d) None of these