Question(1) A successful learner should be able to progress from individual example to broader _____(CO-1)

a. Generalization
b. Inductive reasoning
c. Inductive Inference
d. All the above

Question(2) Select the best learning methods for unrecognized data. (CO-1)

a. Inductive Bias
b. Learning by memorization
c. Inductive reasoning
d. None of the above

Question(3) Learning will be less flexible when we assumed.(CO-1)

a. Strongly
b. Weekly
c. No assumption
d. Both a and b

Question(4) Requirement to use machine learning is(CO-1)

a. Complexity to solve the problem
b. Adaptively to the solution process.
c. Both a and b
d. None of the above

Question(5) The computer program with experience E with respect of some class T and performance measure P, if its performance at task in T, as measured by P, improves with experience E then the process is called.(CO-1)

a. Learning
b. Testing
c. Predicting
d. None of the above

Question(6) Which of them is considered as Task T in the following Checkers problem?(CO-2)

a. Playing Checkers
b. Percentage of games won against the opponents
c. Playing practice game against itself
d. None of the above.

Question(7) To design a learning system, selection of learning is based on (CO-2)

a. The type of training experience from which our system will learn.
b. The training experience is the degree to which the learner controls the sequence of training examples.
c. The training experience is how well it represents the distribution of examples over which the final system performance P must be measured.
d. All of the above

Question(8) The learner might rely on the teacher to select informative board states and to provide the correct move for each in the game of Checker. Which attributes is discussed here for the selection of training experience. (CO-2)

a. Type of training experience
b. The training experience is the degree to which the learners control the sequence of training examples.
c. The training experience is how well it represents the distribution of examples over which the final system performance P must be measured.
d. Non of the above

Question(9) The learner interact with the environment at training time by posing queries or performing experiments.(CO-2)

a. Active learning
b. Passive learning
c. Supervised learning
d. None of the above

Question(10) To minimize the square error E between the training values and the values predicted by the hypothesis $\hat{V}$ as in equation it is required to perform.

$$E = \sum_{(b, v_{train}(b)) \in trainin\ examples} (V_{train}(b) - \hat{V}(b))$$

a. Adjusting the weights
b. Estimating training values
c. Choosing a function approximation algorithm
d. Choosing a representation for the target function

Question(11) To determine exactly what type of knowledge will be learned and how this will be used by the performance program is called.

a. Choosing the target function
b. Adjusting the weights
c. Estimating training values
d. Choosing a function approximation algorithm

Question(12) One key attribute is whether the training experience provides direct or indirect feedback regarding the choices made by the performance system is performed by.

a. Choosing the Training Experience
b. Choosing the target function
c. Adjusting the weights
d. Estimating training values

Question(13) In the game Checker, It takes an instance of a new problem (new game) as input and produces a trace of its solution (game history) as output, the module that must solve the given performance task by using learned target function(s) is called.

a. Performance System

b. Critic
c. Generalizer
d. Experiment Generator

Question(14) In the game of checker, the module takes as input the history or trace of the game and produces as output a set of training examples of the target function is called.

a. Critic
b. Generalizer
c. Experiment Generator
d. Performance System

Question(15) In the game of checker, the module takes as input the training examples and produces an output hypothesis that is its estimate of target function is called.

a. Generalizer
b. Experiment Generator
c. Performance System
d. Critic

Question(16) In the game of checker, the module takes as input the current hypothesis(currently learned function) and output a new problem (i.e. initial board state) for the performance system to explore is called.

a. Experiment Generator
b. Performance System
c. Critic
d. Generalizer

Question(17) The desired output of learning system is.

a. Hypothesis
b. Domain set of the object that we may wish to label.
c. Label set
d. Training data.

Question(18) The training error over the training sample is

$$L_S(h) \stackrel{\text{def}}{=} \frac{|\{i \in [m] : h(x_i) \neq y_i\}|}{m},$$

Where [m] = {1,2, ….m}, h_s is predictor, x_i is input sample and y_i label. This learning paradigm { coming up with a predictor h that minimizes Ls(h) is called

a. Empirical risk minimization
b. True Error
c. Generalization error
d. The risk

Question(19) The predictor whose performance on the training set is excellent, yet its performance on the true “world” is very poor, this phenomenon is called.

a. Overfitting
b. Underfitting
c. No learning

d. None of the above

Question(20) By restricting the learner to choosing a predictor from H(Hypothesis class), we bias it toward a particular set of predictors. Such restrictions are often called.

a. An Inductive bias
b. Overfitting
c. Empirical Error
d. None of the abov

Question(1) Which of the algorithm is not example of decision tree technique?

a. Kmeans
b. ID3
c. CART
d. C4.5

Question(2) The learning method for approximating discrete-valued functions that is robust to noisy data and capable of learning disjunctive expressions is called.

a. Decision tree
b. Regression
c. Support vector machine
d. KNN

Question(3) The tree in decision tree represents.

a. A disjunction of conjunctions of constrains on the attributes of instances
b. A conjunction of disjunctions of constrains on the attributes of instances
c. A conjunction of constrains on the attributes of instances
d. A disjunction of constrains on the attributes of instances

Question(4) The path in the decision tree from root to leaf node is represented as.

a. A conjunction of attribute tests.
b. A disjunction of attribute tests.
c. Both of the above
d. None of the above

Question(5) The root node in ID3 algorithm is identified by selecting parameter which..

a. Classifies the instances better than any of the attributes in the data set
b. Classifies the instances worst than any of the attributes in the data set
c. Do not having specific role in the dataset.
d. None of the above

Question(6) The measure used to select best attributes in ID3 algorithm for classifying data well is known as

a. Information Gain
b. Entropy
c. Split information
d. Information gain ratio

Question(7) Which measure is homogeneity of examples in ID3 algorithm?

a. Entropy
b. Split information
c. Information gain ratio
d. Information Gain

Question(8) Given a collection S, containing positive and negative examples of some target concept, the entropy of S relative to this boolean classification is.

a. $Entropy(S) = -p(+) \log_2 p(+) - p(-)\log_2 p(-)$
b. $Entropy(S) = p(+) \log_2 p(+) + p(-)\log_2 p(-)$
c. $Entropy(S) = p(+) \log_2 p(+) - p(-)\log_2 p(-)$

d. Entropy(S) = $-p(+) \log_2 p(+) + p(-)\log_2 p(-)$

Question(9) If all of the examples in the dataset is positive then entropy is

a. Zero
b. Maximum
c. Average value
d. None of the above

Question(10) If half of the examples in the dataset is positive and remaining are negative then entropy is.

a. Maximum
b. Zero
c. Average value
d. None of the above

Question(11) Measure of expected reduction in entropy is called

a. Information gain
b. Entropy
c. Split Information
d. None of the above

Question(12) On using ID3 algorithm for the samples S with positive samples are 9 and negative samples are 5 and entropy of the dataset is E=0.940 then what is the information gain for then attribute humidity=high has 3 posetive and 4 negative samples and for humidity=normal has 6 posetive and 1 negative sample?

a. 0.151
b. 0.048
c. 0.051
d. 0.121

Question(13) Pure ID3 performs searching decision tree from simple to complex in the hypothesis space using search technique.

a. Hill climbing
b. Depth first search
c. Breadth first search
d. Best first search

Question(14) CART algorithm can problems of

a. Regression and classification
b. Regression only
c. Classification only
d. None of the above

Question(15) Gini Index is used to create decision points for classification tasks in the algorithm

a. CART
b. ID3
c. C4.5
d. C5.0

Question(16) The ID3 algorithm heuristic for slecting the decision tree is

a. Shorter tree over longer ones.
b. Selects trees that place the attributes with highest information gain closest to the root.
c. Both a and b
d. None of the above

Question(17) Approximate Inductive bias of ID3 is the strategy in which.

a. Shorter trees are preferred over larger trees.
b. Trees that place high information gain attributes close to the root are preferred over those that do not.
c. Both a and b
d. None of the them

Question(18) ID3 algorithm decreases its accuracy while increasing the complexity of decision tree when performing on

a. Testing data
b. Training data
c. Both a and b
d. None of the above

Question(19) The use of separate examples, distinct from the training examples, to evaluate the utility of post-pruning nodes of the tree. The given method is applicable to avoid overfiting in decision tree learning by applying the approach.

a. Approaches that stop growing the tree earlier, before it reaches the point where it perfectly classifies the training data.
b. Approaches that allow the tree to overfit the data, and then post-prune the tree.
c. Both a and b
d. None of the above

Question(20) Incorporating continuous value for learning through ID3 algorithm, ID3 need to select the attribute on the basis of

a. Information Gain ratio
b. Information gain
c. Gini Index
d. Split information

Question(1) The approach in which the goal is not learn the underlying distribution but rather to learn an accurate predictor is called.

a. Discriminative
b. Generative
c. Parameter density estimation
d. None of the above.

Question(2) The approach to estimate the specific parametric form over the data for the underlying distribution is known as

a. Parameter density estimation
b. Discriminative method
c. Hypothesis search
d. None of the above

Question(3) The method which provides probabilistic approach to inference.

a. Bayesian reasoning
b. Decision learning method
c. Random forest classification
d. Logistic regression

Question(4) Which of the feature is not included in Bayesian classifier?

a. Bayesian classifier is strict to the observed incremental changes in the training examples.
b. Prior knowledge can be combined with observed data to determine the final probability ~f a hypothesis.
c. Bayesian methods can accommodate hypotheses that make probabilistic predictions.
d. New instances can be classified by combining the predictions of multiple hypotheses, weighted by their probabilities.

Question(5) Practical difficulty in applying Bayesian method is

a. Initial knowledge of many probabilities
b. Not enough data
c. Not enough parameters
d. None of the above

Question(6) The posterior probability P(h|D) where h is hypothesis and D is training data in Bayesian theorem is Inversely proportion to

a. Probability of training data P(D)
b. Prior probability P(h)
c. The P(D|h) to denote the probability of observing data D given some world in which hypothesis h holds.
d. None of the above

Question(7) The argmax P(h/D) where h belongs to H is called a

a. Maximum a posterior hypothesis
b. Posterior probability
c. Prior probability
d. Probability of training data.

Question(8) Consider the city in which 51% are male and 49% are female of the population. The 19% male from male and 2% female from female are job oriented. What is the probability job oriented candidate is male ?

a. S
b. S
c. S
d. S

Question(9) What is argmax value of the function $f(x)=25-12x^2$

a. 0
b. 1
c. Positive Infinite
d. Negative Infinite

Question(10) In the algorithm of Brute force MAP Learning the value of p(h) is where p(h) probability of hypothesis h.

a. 1/H where H is set of Hypothesis
b. 0
c. 1
d. None of the above

Question(11) In Brute force MAP learning algorithm, when h (hypothesis) is inconsistent with D (Training data), the posterior probability is.

a. 0
b. 1
c. 1/H
d. None of the above

Question(12) A normal distribution is characterized by .

a. Mean
b. Standard deviation
c. Both a and b
d. None of the above

Question(13) Bayesian analysis will show that under certain assumptions any learning algorithm that minimizes the squared error between the output hypothesis predictions and the training data will output a

a. Maximum likelihood hypothesis
b. Least likelihood hypothesis
c. Optimal likelihood hypothesis
d. None of the above

Question(14) The limitation of maximum likelihood hypothesis and least squared error hypothesis is

a. Consideration of noise only in target value of the training example.
b. Consider noise in the attributes describing the instances themselves.
c. No consideration of noise.
d. Noise consideration in both a and b

Question(15) In maximum likelihood hypothesis for predicting probabilities the maximum likelihood hypothesis is represented by

a. $h_{ML} = argmax_{\smile} \prod_{i=1}^{m} h(x_i)^{d_i}(1-h(x_i))^{1-d_i}$ $_{h\in H}$

b. $h_{ML} = argmax_{\smile} \prod_{i=1}^{m} h(x_i)^{1-d_i}(1-h(x_i))^{d_i}$ $_{h\in H}$

c. $h_{ML} = argmax_{\smile} \prod_{i=1}^{m} h(x_i)^{1-d_i}(1-h(x_i))^{1-d_i}$ $_{h\in H}$

d. $h_{ML} = argmax_{\smile} \prod_{i=1}^{m} h(x_i)^{d_i}(1-h(x_i))^{d_i}$ $_{h\in H}$

Question(16) The aim to transfer shorter code with minimum bits that is to minimize the expected code length concept proposed by Shannon and Weaver(1949) is used to explain.

a. Minimum Description Length principle
b. Maximum likelihood Hypothesis for predicting probabilities.
c. Least squared error hypothesis.
d. None of the above

Question(17) Minimum description Length principle produces learned tree whose accuracy is comparable to.

a. Standard pruning tree method in decision three
b. ID3 algorithm
c. C4.5 algorithm
d. CART algorithm

Question(18) Given a new instance to classify, the algorithm which simply applies a hypothesis drawn at random according to the current posterior probability distribution is called.

a. Gibbs algorithm
b. Bayes optimal classifier
c. Naïve Bayes classifier
d. C4.5 algorithm

Question(19) The approach to classifying the new instance is to assign the most probable target value, V_{MAP}, given the attribute values ***(a_1 ,a_2 . . .a_n)*** that describe the instance, the approach is called

$$v_{MAP} = argmax_{\smile v_j \in V} P(v_j | a_1, a_2, \ldots . a_n)$$

a. Naïve Bayes classifier
b. Gibbs algorithm
c. ID3
d. Minimum

Question(20) The approach that uses the available observed data of the dataset to estimate the missing data and then using that data to update the values of the parameters.

a. Expectation maximization algorithm
b. Gibbs algorithm
c. Naïve Bayes classifier
d. None of the above

Question(21) The use of Expectation maximization algorithm is

a. It can be used to fill the missing data in a sample.
b. It can be used as the basis of unsupervised learning of clusters.

c. It can be used for discovering the values of latent variables.
d. All of the above