

- 001.** Which of the following is subset of Machine Learning **C**  
 A Numpy B Pandas  
 C Deep Learning D Sklearn
- 002.** What kind of learning algorithm for Future stock prices or currency exchange rates? **B**  
 A Recognizing anomalies B Prediction  
 C Generating patterns D Recognizing patterns
- 003.** What is Machine Learning? **D**  
 A The selective acquisition of knowledge through the use of manual programs B The selective acquisition of knowledge through the use of computer programs  
 C The autonomous acquisition of knowledge through the use of manual programs D The autonomous acquisition of knowledge through the use of computer programs
- 004.** Artificial Intelligence is about\_\_\_\_\_ **B**  
 A Playing a game on Computer B Making a machine Intelligent  
 C Programming on Machine with your Own Intelligence D Putting your intelligence in Machine
- 005.** Who is known as the -Father of AI"? **C**  
 A Fisher Ada B Alan Turing  
 C John McCarthy D Allen Newell
- 006.** Identify the model which is trained with data in only a single batch **C**  
 A Online learning B Offline learning  
 C Batch learning D Group learning
- 007.** Identify the type of learning in which labeled training data is used **A**  
 A Supervised learning B Unsupervised learning  
 C Reinforcement learning D Semi unsupervised learning
- 008.** Identify the kind of learning algorithm for facial identities for facial expressions **B**  
 A Predictions B Recognition patterns  
 C Recognizing anomalies D Generating patterns
- 009.** Fraud detection are application of **B**  
 A Unsupervised learning: clustering B Supervised learning: classification  
 C Reinforcement Learning D Unsupervised learning: Regression
- 010.** Among the following options identify the one which is false regarding regression **D**  
 A It is used for the prediction B It is used for the interpretation  
 C It relates inputs to outputs D It discovers casual relationships
- 011.** What is unsupervised learning? **C**  
 A Number of groups may be known B Features of the groups explicitly stated  
 C Neither features nor number of groups known D It has labeled data
- 012.** The father of machine learning is \_\_\_\_\_ **A**  
 A Geoffrey Everest Hinton B Geoffrey Hill  
 C Geoffrey Chaucer D Geoffrey Ritchie
- 013.** What is the term known as on which the machine learning algorithms build a model based on sample data? **B**  
 A Data training B Training data  
 C Transfer data D Testing data
- 014.** Machine learning is a subset of which of the following **A**  
 A Artificial intelligence B Deep learning  
 C Data learning D Online learning
- 015.** In \_\_\_\_\_ learning, the training data is unlabeled **B**  
 A Supervised B Unsupervised  
 C Reinforcement D Semi supervised
- 016.** Which of the following is not an unsupervised learning algorithm **D**

- A k-Means B Hierarchical Cluster Analysis  
C Apriori D Classification

**017.** What is the most common issue when using Machine Learning? **A**  
A Poor Data Quality B Lack of skilled resources  
C Inadequate Infrastructure D Cost of Software

**018.** What is another name for an input attribute? **D**  
A Predictive variable B Estimated variable  
C Dependent variable D Independent variable

**019.** What is the output of training process in machine learning? **C**  
A Null B Accuracy  
C Machine learning model D Machine learning algorithm

**020.** In supervised learning \_\_\_\_\_ **B**  
A Classes are not predefined B Classes are predefined  
C Classes are not required D Classification is not done

**021.** Which of the following are categorical features? **C**  
A Height of a person B Price of petroleum  
C Mother tongue of person D Amount of rainfall in a day

**022.** Which of the following is not a method to handle missing or corrupted data in a dataset? **D**  
A Drop missing rows or columns B Assign a unique category to missing values  
C Replace missing values with mean/median/mode D Change the feature name

**023.** In statistics \_\_\_\_\_ is the entire set of items from which you draw data for a statistical study **A**  
A Population B Sampling  
C Simpling D Dataset

**024.** The \_\_\_\_\_ error is an error from erroneous assumptions in the learning algorithm **A**  
A Bias B Variance  
C Accuracy D Precision

**025.** \_\_\_\_\_ processes the uncategorized data and divides them into different clusters. **A**  
A Clustering algorithm B Regression algorithm  
C Classification algorithm D Reinforcement algorithm

**026.** Machine learning gives computers the ability to learn without being explicitly programmed said by \_\_\_\_\_ **A**  
A Arthur Samuel B Tom Mitchell  
C Alan Turing D Mc. Karthi

**027.** Branch of Engineering student is a \_\_\_\_\_ type feature **C**  
A Continuous B Ordinal  
C Nominal D String

**028.** In which of the following type of learning the teacher returns reward and punishment to the learner? **B**  
A Active learning B Reinforcement learning  
C Supervised learning D Unsupervised learning

**029.** Which of the following is unsupervised task? **A**  
A Grouping images of footwear and caps separately for a given set of images  
B Learning to play chess  
C Predicting if an edible item is sweet or spicy based on the information of the ingredients and their quantities  
D Prediction of house pricing

**030.** In \_\_\_\_\_ learning, you train the system incrementally by feeding it data instances sequentially, either individually or by small groups. **A**

- A Online B Offline  
C Batch D Single

031. \_\_\_\_\_ occurs when data is unable to establish an accurate relationship between input and output variables A  
A Under fitting B Over fitting  
C Best fitting D General fitting

032. Which of the following is not a category in unsupervised learning D  
A Clustering B Visualization and dimensionality reduction  
C Association rule learning D Regression

033. What is the application of machine learning methods to a large database called? C  
A Big data computing B Internet of things  
C Data mining D Artificial intelligence

034. Which of the following statement is true about prediction problems? D  
A The output attribute must be numeric. B The output attribute must be categorical  
C The resultant model is designed to determine future outcomes D The resultant model is designed to classify current behavior

035. The frequency distribution of individual data points in the original dataset is called A  
A Data distribution B Data plotting  
C Sampling D Visualizing

036. Which of the following does not include different learning methods? B  
A Analogy B Introduction  
C Memorization D Deduction

037. The \_\_\_\_\_ is an error from sensitivity to small fluctuations in the training set B  
A Bias B Variance  
C MSE D RMSE

038. Under fitting can be tackled by using \_\_\_\_\_ D  
A Analyzing the data with the utmost level of perfection B Use data augmentation technique  
C Remove outliers in the training set D Maximize the training time

039. A sample is defined as a smaller and more manageable representation of a larger group B  
A Population B Sampling  
C Simpling D Dataset

040. \_\_\_\_\_ is used for visualization of data distribution A  
A Histogram B Barplot  
C Scatterplot D Heatmap

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

042. Which of the following is incorrect D  
A High model complexity tends to have a low bias B High model complexity tends to have a high variance  
C High bias may cause to underfitting D Low variance may cause to overfitting

043. In \_\_\_\_\_ learning there are normally no parameters to tune, the system is normally hard-coded with priors in the form of fixed weights A  
A Instance-based B Model-based  
C Online D Batch

044. Overfitting can be tackled by using \_\_\_\_\_ D  
A Enhance the complexity of the model B Add more features to the data  
C Reduce regular parameters D Select a model with lesser features

045. Which of the following is a supervised learning problem? i) Predicting the outcome of a D

cricket match as a win or loss based on historical data ii) Recommending a movie to an existing user on a website like IMDB based on the search history iii) Predicting the gender of a person iv) Predicting the classes of articles

A I, II, III

B I, III, IV

C II, III, IV

D I, II, III, IV

**046.** Which of the followings are classification tasks? i) Find the gender of a person by analyzing his writing style ii) Predicting the price of a house based on the floor area, the number of rooms iii) Predict whether there will be abnormally heavy rainfall next year iv) Predict the number of copies of a book that will be sold this month **C**

A I, II

B II, III, IV

C I, III

D I, III, IV

**047.** What is Machine Learning? i) Artificial Intelligence ii) Deep Learning iii) Data Statistics **C**

A Only i

B Only ii

C i and ii

D i and iii

**048.** Regression algorithms are used to predict the \_\_\_\_\_ values **A**

A Continuous

B Discrete

C Categorical

D Continuous and categorical

**049.** Which supervised learning technique can process both numeric and categorical input attributes? **B**

A Bayes classifier

B Linear regression

C Logistic regression

D Support vector classifier

**050.** Logistic regression is a \_\_\_\_\_ technique **B**

A Regression

B Classification

C Clustering

D Bagging

**051.** Missing data items are ..... with Bayes classifier **C**

A Ignored

B Treated as equal compares

C Treated as unequal compares.

D Replaced with a default value.

**052.** Which of the following is not a supervised learning? **B**

A Naive Bayesian

B PCA

C Linear Regression

D Decision Tree

**053.** SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called **B**

A support machine

B support vectors

C support points

D support line

**054.** Which of the following is not a valid SVM type? **C**

A Linear SVM

B Non-linear SVM

C Multiple SVM

D Kernel SVM

**055.** Data used to optimize the parameter settings of a supervised learner model is called .....? **C**

A Test

B Training

C Validation

D Verification

**056.** Regression trees are often used to model which data? **A**

A Linear

B Nonlinear

C Categorical

D Ordinal

**057.** What is called the average squared difference between classifier predicted output and actual output? **B**

A Mean relative error

B Mean squared error

C Mean absolute error

D Root mean squared error

**058.** \_\_\_\_\_ defines how far the line is shifted during each step, based on the information from the previous training step **C**

A Training rate

B Testing rate

C Learning rate

D Predicting rate

**059.** \_\_\_\_\_ is an extreme value that greatly differs from the other values **B**

A Missing values

B Outliers

- C Non scaled values                      D Dummy values
- 060.** Random forest is a well-known machine learning algorithm that uses.. **A**  
 A Supervised learning                      B Unsupervised learning  
 C Hybrid learning                      D Semi-supervised learning
- 061.** Which of the following is not a type of naive bayes model? **D**  
 A Gaussian                      B Multinomial  
 C Bernoulli                      D Polynomial
- 062.** A regression model in which more than one independent variable is used to predict the dependent variable is called . **B**  
 A A simple linear regression                      B A multiple regression  
 C An independent model                      D A dependent model
- 063.** \_\_\_\_\_ is used to minimize the MSE by minimizing the cost function value in linear regression **A**  
 A Gradient descent                      B Euclidian  
 C Bernoulli                      D Elbow method
- 064.** Explained variation/total variation is a formula for **D**  
 A MSE                      B RMSE  
 C R Score                      D R2 Score
- 065.** The distance between the actual value and predicted values is called \_\_\_\_\_ **C**  
 A Outliers                      B Anomalies  
 C Residuals                      D Bias
- 066.** \_\_\_\_\_ in regression analysis occurs when two or more independent variables are closely related to each other **D**  
 A Under-fitting                      B Over-fitting  
 C Appropriate-fitting                      D Multicollinearity
- 067.** High variance is caused by \_\_\_\_\_ **D**  
 A Under-fitting                      B Over-fitting  
 C Appropriate-fitting                      D Multicollinearity
- 068.** \_\_\_\_\_ is a metric to measure the impurity in a given attribute, used in decision tree algorithm. **B**  
 A pruning                      B Entropy  
 C Mean                      D Standard deviation
- 069.** Which of the following algorithm works based on ensemble learning? **D**  
 A Linear regression                      B SVM  
 C KNN                      D Random forest
- 070.** While implementing a Decision tree, the main issue arises that how to select the best attribute for the root node and for sub-nodes. So, to solve such problems there is a technique which is called as ASM, stands for \_\_\_\_\_ **A**  
 A Attribute selection measure                      B Automatic selection measure  
 C Attribute separation measure                      D Automatic separation measure
- 071.** In order to build a tree in decision tree algorithm, we use the CART algorithm, which stands for \_\_\_\_\_ **A**  
 A Classification and Regression Tree algorithm                      B Continuous and Regression Tree algorithm  
 C Classification and Rooted Tree algorithm                      D Continuous and Rooted Tree algorithm
- 072.** \_\_\_\_\_ is the process of removing the unwanted branches from the tree. **B**  
 A Shrinking                      B Pruning  
 C Dropping                      D Truncate
- 073.** The \_\_\_\_\_ matrix is a matrix used to determine the performance of the classification models for a given set of test data **A**  
 A Confusion                      B Creative  
 C Correlation                      D Regression
- 074.** Model has given prediction No, and the real or actual value was also No. then it is **B**

termed as \_\_\_\_\_

- |                  |                  |
|------------------|------------------|
| A True positive  | B True negative  |
| C False positive | D False negative |

- 075.** Which shape of graph we can find in logistic algorithm? **B**  
A L B S  
C R D U
- 076.** Which of the following algorithm gives the probabilistic values which lie between 0 and 1. **D**  
A SVM B KNN  
C Linear Regression D Logistic Regression
- 077.** In logistic regression which is used as a cost function? **D**  
A MSE B RMSE  
C MAE D Sigmoid function
- 078.** The Bayes rule can be used in ..... **D**  
A Solving queries B Increasing complexity  
C Decreasing complexity D Answering probabilistic query
- 079.** Choose a disadvantage of decision trees among the following. **C**  
A Decision trees are robust to outliers B Factor analysis  
C Decision trees are prone to be overfit D Decision trees are prone to be underfit
- 080.** Among the following identify the one in which dimensionality reduction reduces. **D**  
A Performance B Entropy  
C Stochastics D Collinearity
- 081.** Which of the following machine learning algorithm is based upon the idea of bagging? **B**  
A Decision tree B Random forest  
C Classification D Regression
- 082.** Which of the following is not an example of Naive Bayes Algorithm **D**  
A Spam filtration, B Sentimental analysis  
C Classifying articles. D Customer segmentation
- 083.** \_\_\_\_\_ algorithm stores all available data and classifies a new data point based on its similarity to the existing data. **A**  
A Naive Bayes B KNN  
C Decision tree D SVM
- 084.** In SVM we need to find out the best decision boundary that helps to classify the data points. This best boundary is known **B**  
A Hyperbola B Hyperplane  
C Gaussian boundary D Elbow boundary
- 085.** KNN algorithm is also called . **A**  
A lazy learner algorithm B Easy learner algorithm  
C Crazy learner algorithm D Middle level algorithm
- 086.** In Decision Tree, Decision Nodes are represented by \_\_\_\_\_ **B**  
A Disks B Squares  
C Circles D Triangles
- 087.** Which of the following machine learning algorithms has both training and test phases? **B**  
A k-Nearest Neighbor B Linear regression  
C Case-based reasoning D All machine learning algorithms
- 088.** Given a kNN classifier, which one of the following statements is true? **B**  
A The more examples are used for classifying an example, the higher accuracy we obtain B The more attributes we use to describe the examples the more difficult is to obtain high accuracy  
C The costliest part of this method is to learn the model D We can use KNN for classification only
- 089.** What is the way to ensemble multiple classifications or regression? **D**  
A Bagging B Blending

## C Boosting

## D Stacking

**090.** 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 **D**

A i and ii

B      ii and iii

C    iii and iv

D i, ii and iii

**091.** Logistic regression is a ..... regression technique that is used to model data having a ..... outcome. **C**

A Linear, binary

B Linear, numeric

C Nonlinear, binary

D Nonlinear, numeric

**092.** Accuracy is one of the important parameters to determine the accuracy of the classification problems. The formula used to find it is **A**

A  $(TP+TN) / (TP+TN+FP+FN)$

B  $(TP+FN) / (TP+TN+FP+FN)$

C  $(FP+TN) / (TP+TN+FP+FN)$

D  $(FP+FN) / (TP+TN+FP+FN)$

**093.** The \_\_\_\_\_ is a graph displaying a classifier's performance for all possible thresholds. The graph is plotted between the true positive rate (on the Y-axis) and the false Positive rate (on the x-axis). **B**

A NOC

B ROC

### C Counting plot

D Scatter plot

**094.** Entropy(S)-[(WeightedAvg)\*Entropy (eachfeature)], is a formula used in decision tree algorithm to find out ... **A**

A Information gain

B Gini index

C Depth of the tree

D Height of the tree

**095.** Mathematically, we can represent a linear regression as  $y = a_0 + a_1x + \epsilon$ . Here  $\epsilon$  indicates **D**

A Dependent Variable

B Independent Variable

C    intercept of the line

D random error