**MCQ**

1. Which of the metrics is suited for evaluating any classification algorithm?
   1. Mean Squared Error
   2. Mean Absolute Error
   3. Log Loss
   4. Mean Squared Logarithmic Error

Ans: c

1. Precision is ratio of
   1. True Positive / Total number of Positives in actual (test) data set
   2. True Positive / Total number of Positives among the predicted values
   3. True Positive / Total number of Positives and Negatives in actual (test) data set
   4. True Positive / Total number of Positives and Negatives among the predicted values

Ans: d

1. Given a case problem of classifying images, which of the metric can be the best(among the given below) for optimizing the accuracy of any deep learning model used?
   1. Precision
   2. Mean Absolute Error
   3. Negative Log Loss
   4. Mean Squared Error

Ans: c

1. Which of the algorithms can be used only for classification?
   1. Naïve Bayes
   2. K-NN
   3. Decision Trees
   4. SVM

Ans: a

1. In L1 regularization, the error function involves
   1. Sum of Squared values of all the weights/coefficients
   2. Sum of Absolute values of all the weights/coefficients
   3. Sum of Square root of squared values of all the weights/coefficients
   4. Simply sum of all the weights/coefficients

Ans: b

1. The gradient descent method, if the learning rate is kept very high then,
   1. Minima of the error might not be reached
   2. Maxima of the error will be reached
   3. Minima of the error surely will be reached
   4. Maxima of the error will not be ever reached

Ans: a

1. When we tune the model on a data with function like gridsearchCV (sklearn), then
   1. The tuning happens on some randomly chosen set of parameters with K-Fold cross-validation
   2. The tuning happens on all the possible parameters set of parameters specified with just a 2 partitioned train-test split
   3. The tuning happens on all the possible set of parameters specified with K-Fold cross-validation
   4. The tuning happens on all the possible parameters set of parameters specified with K-Fold cross-validation

Ans: c

1. When we tune the model on a data with function like randomizedsearchCV (sklearn), then
   1. The tuning happens on some randomly chosen set of parameters with K-Fold cross-validation
   2. The tuning happens on all the possible parameters set of parameters specified with just a 2 partitioned train-test split
   3. The tuning happens on all the possible set of parameters specified with K-Fold cross-validation
   4. The tuning happens on all the possible parameters set of parameters specified with K-Fold cross-validation

Ans: d

1. Which of the following algorithms assume the multivariate normal distribution of the data?
   1. Gaussian Naïve Bayes
   2. Support Vector Machines
   3. Discriminant Analysis
   4. Decision Trees

Ans: c

1. In support vector machines, the radial basis function has which tuning parameters mainly?
   1. Cost and gamma
   2. Cost only
   3. Cost and degree
   4. Cost and coef0

Ans: a

1. When we have multi-class classification case for neural network, which should be the activation for the output layer?
   1. Softmax
   2. ReLU
   3. Sigmoid
   4. Tanh

Ans: a

1. When we have regression case for neural network, which should be the activation for the output layer?
   1. Softmax
   2. ReLU
   3. Sigmoid
   4. Tanh

Ans: b

1. Suppose in a decision tree classifier at one internal node with gini=0.67, there are two four possible splits(mentioned in options) for a data with some to features X1, X2,…,X10. Which of the splits will be preferred?
   1. X1 <= 34.4 with reduction in gini as 0.3
   2. X9 <= 6.7 with reduction in gini as 0.43
   3. X4 <= 9.3 with reduction in gini as 0.1
   4. X1 <= 10.9 with reduction in gini as 0.28

Ans: b

1. Suppose in a decision tree classifier at one internal node has prevalence of two classes A and B as 0.4 and 0.6 respectively. What can be the Gini’s Impurity Index calculated?
   1. 0.48
   2. 0.52
   3. 0.24
   4. 0.76

Ans: a

1. Bagging technique implementation in scikit-learn uses which of the estimator as the default base estimator?
   1. Support Vector Machines
   2. Random Forest
   3. K-NN
   4. Decision Trees

Ans: d

1. Bootstrap Sampling is involved in which of the techniques?
   1. XG Boosting
   2. Linear Regression
   3. Neural Networks
   4. Random Forest

Ans: d

1. What is not true about h2o technology?
   1. H2o is open source technology
   2. H2o is compatible with R
   3. H2o has a function that directly implements association rules mining
   4. H2o has a function that directly implements k-means clustering

Ans: c

1. On your machine, h2o engine runs on
   1. Java Virtual Machine
   2. Apache Tomcat
   3. SQL server
   4. Web Server

Ans: a

1. Logistic regression makes use of
   1. Sigmoid function
   2. Tanh function
   3. ReLU function
   4. CV function

Ans: a

1. For Linear Discriminant Analysis, it is assumed that,
   1. The variations for all the categories are same
   2. The means for all the categories are same
   3. The variations for all the categories are different
   4. The means for all the categories are different

Ans: a

1. In boosting classifier,
   1. The algorithm is applied multiple times to the same bootstrap sampled data
   2. The algorithm is applied to the data on which more weights are applied for wrongly classified observations and processed repeated
   3. The algorithm is applied only once to different bootstrap sampled datasets
   4. The algorithm is applied to a data just once and it is done

Ans: b

1. Which of these are the tuning parameter set for gradient boosting algorithm?
   1. Number of trees, learning rate, number of splits
   2. Gamma, cost, learning rate
   3. Number of trees, number of splits, nearest neigbours
   4. Gamma, nearest neighbour

Ans: a

1. Which of the ML algorithms can be involved in voting classifier?
   1. Logistic Regression, Linear Regression, K-NN classifier
   2. K-NN classifier, SVM classifier, Logistic Regression
   3. SVM regressor, Linear Regression, MLP classifier
   4. Linear regressor, MLP regressor, SGD regressor

Ans: b

1. Which of the ML algorithms can be involved in voting regressor?
   1. Logistic Regression, Linear Regression, K-NN classifier
   2. K-NN classifier, SVM classifier, Logistic Regression
   3. SVM regressor, Linear Regression, MLP classifier
   4. Linear regressor, MLP regressor, SGD regressor

Ans: d

1. R squared value can be at most,
   1. 0
   2. 0.5
   3. 1
   4. -1

Ans: c

1. In Hierarchical clustering, single linkage refers to
   1. Considering distance any single observation from cluster with any single observation in another cluster
   2. Considering distance between two farthest observations of two different clusters
   3. Considering distance between to nearest observations of two different clusters
   4. Considering the average distance between the two clusters

Ans: c

1. In Hierarchical clustering, complete linkage refers to
   1. Considering distance any single observation from cluster with any single observation in another cluster
   2. Considering distance between two farthest observations of two different clusters
   3. Considering distance between to nearest observations of two different clusters
   4. Considering the average distance between the two clusters

Ans: b

1. In K-Means clustering, the criterion of choosing the best number of clusters can be,
   1. Within Sums of squares
   2. R squared
   3. ROC AUC
   4. Accuracy

Ans: a

1. In K-Means clustering, at first centroid (1st iteration) are chosen,
   1. At random
   2. Calculated as observations with minimum value
   3. Calculated as observations with maximum value
   4. Calculated as observations with average value

Ans: a

1. It is necessary to scale the data because,
   1. There may be variables with different data types in the data
   2. There may be variables with different magnitudes in the data
   3. There may be variables with same name in the data
   4. There may be variables with same mean in the data

Ans: b

1. Min-Max scaler can enable us to scale the data
   1. To the data with same mean
   2. To the data with same variance
   3. To the data with range between mean-sd and mean+sd
   4. To the data with range between 0 to 1

Ans: d

1. Consider that there is a rule if{A,F,G} then {W,P} with Confidence as 0.8 and proportion(support) of transactions of items{W,P} are 0.9. What can you comment on this rule?
   1. When items {A,F,G} are there then items {W,P} are also there
   2. Items {W,P}, if they are items to be sold then those should be kept near the items {A,F,G}
   3. The rule should not be considered because the lift ratio of the rule is less than 1, whatever is the confidence
   4. The rule should be highly considered because the lift ratio of the rule is greater than 1, with high confidence

Ans: c

1. If a term called “evolution” occurs in a document 15 times which contains 300 terms than term frequency of the term “evolution” for that document will be
   1. 15/300
   2. 15 x 300
   3. log(300/15)
   4. 300/15

Ans: a

1. If the term “liberty” occurs in 3000 documents among the corpus of 15000 documents then inverse document frequency of the term “liberty” can be calculated as
   1. 15000/3000
   2. log(15000/3000)
   3. 3000/15000
   4. 3000 x 15000

Ans: b

1. Moving Model (MA model) is model considering
   1. Dependence of today’s value with yesterday’s value or before
   2. Dependence of today’s value with tomorrow’s value
   3. Dependence of today’s value with yesterday’s errors or before
   4. Dependence of today’s value with average of all previous values

Ans: c

1. White noise has
   1. Fixed mean and fixed variance
   2. Normal distribution
   3. No specific mean and variance
   4. Fixed mean and varying variance

Ans: a

1. Holt-Winters Method
   1. Considers smoothing over the period of time
   2. Considers trend as well as seasonality
   3. Considers only seasonality
   4. Considers only trend

Ans: b

1. Cyclical component in time series decomposition signifies
   1. Rise and fall at regular time intervals
   2. Rise at regular time intervals
   3. Fall at regular time intervals
   4. Rise and Fall not at regular time intervals

Ans: d

1. Auto-encoders resemble which of the unsupervised learning algorithms use-wise?
   1. Clustering analysis
   2. Principal Component Analysis
   3. Association Rules Mining
   4. Support Vector Machines

Ans: b

1. LSTM model is a modification of
   1. Convolutional Neural Network
   2. Auto-encoders
   3. Bellman Equation
   4. Recurrent Neural Network

Ans: d