

NPTEL Course: Business Analytics and Data Mining Modelling Using R

Week 1: Assignment 1

1. Classify each of the following as N (nominal), O (ordinal), or I/R (interval/ratio) data:

Pin code of a city, quality of clothes you have, place of your study, last CGPA you achieved in exam, mileage of a car

- A. I/R, O, N, I/R, I/R.
- B. N, I/R, N, O, O.
- C. N, O, N, I/R, I/R.
- D. I/R, N, O, I/R, N.

2. If two variables have a correlation coefficient value of 0.01. What does it suggest?

- A. It suggests positive correlation.
- B. It suggests causality.
- C. It suggests no correlation.
- D. None of the above.

3. Which type of analytics mostly employs machine learning techniques?

- A. Decision making.
- B. Prescriptive.
- C. Descriptive.
- D. Predictive.

4. Which scenarios might create overfitting issues?

- A. By training a model using a complex function that fits the data perfectly
- B. By training a model that ends up fitting the noise and explaining the chance variation
- C. By training a model with more no. of iterations resulting in excessive learning of the data
- D. All of the above.

5. Which of the following is not true about unsupervised learning?

- A. It might involve clustering of the data.
- B. Finding interesting patterns hidden in the data.
- C. Same as semi-supervised learning.
- D. None of the above.

6. Surveying all individuals of a given population is referred to as

- A. Sampling
- B. Poll
- C. Census
- D. None of the above

7. Which one of the following is the benefit of using simple random sampling?

- A. Informants won't refuse to participate.
- B. Interviewers can choose respondent freely.
- C. The results are always representative.
- D. None of the above.

8. Which of the following are true about dummy coding?
- A. Dummy binary variables having 0's and 1's: 0 indicates 'absence' and 1 indicates 'presence'
 - B. Dummy binary variables having 0's and 1's: 0 indicates 'presence' and 1 indicates 'absence'
 - C. Used for categorical variables
 - D. None of the above.
9. Which of the following are true assumptions made in Student's t-test?
- A. The underlying population distributions have equal variance.
 - B. The underlying population follows a non-symmetrical distribution.
 - C. The underlying population follows normal distribution.
 - D. All of the above.
10. Which function is used to print all the variable names in a data frame df in R?
- A. names()
 - B. names(df)
 - C. df.names()
 - D. names("df")

Course: Business analytics and data mining Modelling using R

Week-2, Assignment-2

1. Which of the following is true about data and data preprocessing?
 - a) Raw data is the data obtained after executing pre-processing steps.
 - b) Pre-processed data is the original form of data.
 - c) Raw data is the main source of data.
 - d) None of the above.
2. Which of the following statement is wrong about dataset partitioning?
 - a) Partitioning helps in addressing bias in the model.
 - b) Partitioning creates multiple subsets of data which are then used for data visualisation.
 - c) Partitioning can be used to create a validation set which is then used to finetune and improve the model.
 - d) None of the above.
3. Which of the following statements indicate the true advantages of using panel data over pure cross-sectional or pure time series data?
 - I: Panel data can model both the common and individual differences in the population.
 - II: Panel data packs more information and variability than pure cross-sectional or pure time series data
 - III: Panel data allows the analyst to estimate fewer statistical effects than pure cross-sectional or pure time series data.
 - a) I only.
 - b) I and II only.
 - c) II only.
 - d) None of the above.
4. Which of the following data is the outcome of data cleaning and data preparation steps?
 - a) Raw data.
 - b) Pre-processed data.
 - c) Synchronised data.
 - d) None of the above.
5. When a dependent variable is restricted to a certain range of values while the independent variables are allowed to have their corresponding values, this step is most accurately described as:
 - a) Censoring.
 - b) Truncation.
 - c) Discretization
 - d) None of the above.

6. Which of the following steps are executed in the data exploration phase?
- a) Plot various data visualisations to shortlist the data mining techniques to be used.
 - b) Free-form exploration of data to help the formal analyses.
 - c) Identify bin sizes to convert continuous variables into categorical variables.
 - d) None of the above.
7. Which of the followings is best suited to compare the distributions of more than one group in the data?
- a) Box plots.
 - b) Histogram.
 - c) Line graph.
 - d) None of the above
8. Which of the following statements are not the most accurate descriptions of data visualisation?
- a) Box plots are useful to understand the distribution of the data and missing values.
 - b) Bar graphs are suitable for time series data.
 - c) Line graphs are not suitable for time series data.
 - d) None of the above
9. What does a large standard deviation suggest?
- a) The values are widely distributed and the mean may not be a reliable measure of central tendency.
 - b) The values are widely distributed and the median would be an unreliable measure of the central tendency.
 - c) Values are not normally distributed.
 - d) None of the above.
10. Histograms, line charts, and scatter plots are all types of:
- a) 2D diagram.
 - b) 1D diagram
 - c) 3D diagram.
 - d) None of the above.

NPTEL Course: Business analytics and data mining Modeling using R

Week-3, Assignment-3

1 Which of the followings are not true about heatmaps?

- a) Useful to visualize correlations between variables.
- b) Useful to visualize class separation.
- c) Color themes are used to indicate values.
- d) None of the above.

2 Which of the followings are true about multi-dimensional visualisation?

- a) Multiple panels having plots between different variables.
- b) Many plots each with one or two variables.
- c) Marker color can be used to represent 3rd variable.
- d) None of the above.

3 Which of the following steps would be better for reducing dimensionality of a dataset?

- a) Removing columns which have too many missing values.
- b) Removing columns with dissimilar data trends.
- c) Removing columns which have high variance in data.
- d) None of the above.

4 Which of the following is true about principal component analysis (PCA)?

- I : PCA is a supervised technique.
- II : PCA searches for the directions where data have the largest variance.
- III : Maximum number of principal components is equal to number of features.
- IV : All principal components are orthogonal to each other.

- a) I and II
- b) I, II, and IV
- c) II and III
- d) None of the above

5 Which of the followings are dimension reduction techniques?

- I: Domain knowledge
- II: Data exploration techniques
- III: Data conversion techniques
- IV: Automated reduction techniques
- V: Data mining techniques

- a) II, IV and V
- b) I, III and IV
- c) I, IV and V
- d) None of the above

6 When we add an irrelevant feature to a linear regression model, it may result in:

I : An increase in R-square

II : A decrease in R-square

- a) Only I is correct
- b) Only II is correct
- c) Either I or II is correct
- d) None of the above

7 Which of the following is suitable only for predicting a class?

- a) Linear regression
- b) Correlation
- c) Decision trees
- d) Naïve Bayes

8 Which of the following is true about correlation analysis?

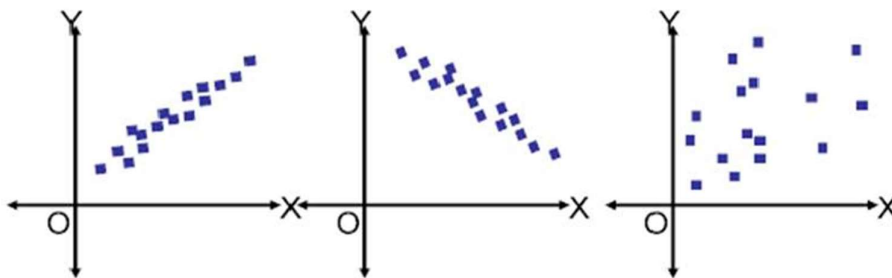
- a) Used to estimate the effects of multiple independent variables on a dependent variable.
- b) Used to predict values of y based on values of x .
- c) Used to measure the strength of association between two variables
- d) None of the above

9 Which of the following is a way to measure the degree of linear association between two quantitative variables?

- a) Pearson correlation coefficient
- b) Significance level
- c) p-value
- d) None of the above

10 Consider the three scatter plots given below, which of the followings are examples of multi-collinear features?

- a) Features in plots 1&3
- b) Features in plots 2&3
- c) Features in plots 1&2
- d) None of the above



NPTEL Course: Business analytics and data mining Modeling using R

Week - 4, Assignment - 4

- Which of the following metrics measure the 'goodness of fit' of a regression model?
 - Mean absolute deviation
 - Root mean squared error
 - R-square
 - Total sum of squared errors
- Which of the following statements is true about outliers in all scenarios?
 - Outliers should be identified and removed from a dataset.
 - Outliers should be part of the training set but should not be present in the test set.
 - Outliers should be part of the test set but should not be present in the training dataset.
 - Nature of the problem determines how to deal with the outliers.
- Which of the following is the average positive difference between estimated and actual outcome values?
 - Root mean squared error
 - Mean squared error
 - Mean absolute error
 - None of the above
- Which of the followings are common uses of unsupervised techniques like clustering?
 - Detect outliers
 - Find a relevant set of input features for supervised learning
 - Find meaningful relationships in the dataset
 - None of the above.
- Consider a classifier for screening of terrorists in a military application (Positive means action has to be taken). The confusion matrix is shown below.

		Predicted	
		Positive	Negative
Actual	Positive	TP	FN
	Negative	FP	TN

Which of the following situations would be acceptable if your classifier has it?

- FP >> FN
 - FN >> FP
 - TN >> FP
 - None of the above
- Which of the following are accurate axes labels for an ROC curve?
 - Vertical axis: % of true negatives; Horizontal axis: % of false negatives
 - Vertical axis: % of true positives; Horizontal axis: % of false positives
 - Vertical axis: % of false negatives; Horizontal axis: % of false positives
 - Vertical axis: % of false positives; Horizontal axis: % of true negatives

7. Which of the following are correct scenarios to use oversampling?
- a) When the cost of failing rare events are low.
 - b) To de-emphasize rare events to the learning algorithm
 - c) When a binary response variable includes significantly more items with one choice of response than the other choice, and we seek to accurately predict both the choices.
 - d) When a binary response variable includes roughly the same number of items for each choice.
8. Which is not included as a step when partitioning is done with oversampling.
- a) Half the records from class 1 stratum are randomly selected into training partition.
 - b) Randomly select class 0 records for training partition equal to no. of class 1 records.
 - c) Randomly select class 0 records to maintain the original ratio of class 0 to class 1 records for validation partition.
 - d) None of the above.
9. Which of the following statements is true about prediction problems in general?
- a) The output attribute must be categorical.
 - b) The output attribute must be numeric.
 - c) The resultant model is designed to determine future outcomes.
 - d) The resultant model is designed to classify current behavior.
10. Which of the following are true about asymmetric misclassification cost?
- a) Considered when the misclassification error for a class of interest is more costly than for the other class.
 - b) Goal of performance metrics changes to either minimization of costs or maximisation of profits.
 - c) Performance metrics which incorporate asymmetric misclassification costs are used.
 - d) None of the above.

NPTEL Course: Business analytics and data mining Modeling using R

Week - 5, Assignment – 5

1. Which of the following is correct about prediction performance?
- a) The focus is on reducing root-mean-square error of the continuous output variable
 - b) The focus is on reducing misclassification error of the continuous output variable
 - c) Prediction performance is same as classification performance
 - d) None of the above

2. Which of the following is true about the statements below given?

I: 'Goodness of fit' focuses on how well the model fits the data

II: The model obtained in data mining is certainly same as the one obtained in statistical setting

- a) Only statement I is correct
 - b) Only statement II is correct
 - c) Both the statements are correct
 - d) None of the above
3. Which partition is used for the performance evaluation of the data mining model?
- a) Training partition
 - b) Validation partition
 - c) Test partition
 - d) None of the above
4. What does the negative overall average error indicate?
- a) On an average level, the model is over predicting the observations
 - b) On an average level, the model is under predicting the observations
 - c) On an average level, the model has no impact on predicting the observations
 - d) None of the above

5. Consider the following statements and choose the correct option:

Assertion: RMSE is one of the most preferred metrics of predictive accuracy

Reasoning: It is computed in the same unit as the outcome variable and similar to the standard error of estimate

- a) Assertion is correct and reasoning gives the correct explanation
 - b) Assertion is correct but reasoning does not give the correct explanation
 - c) Assertion is incorrect but reasoning is correct
 - d) Both assertion and reasoning are incorrect
6. Which of the following is true about predictive modelling using multiple linear regression?
- a) Used to estimate best-fit model
 - b) Used to estimate values of outcome variable for new records
 - c) Prediction performance is of secondary importance
 - d) None of the above

7. Which of the following is correct about OLS?

- a) OLS computes the sample estimates which minimises the sum of deviations between predicted and actual values
- b) OLS computes the sample estimates which maximises the sum of deviations between predicted and actual values
- c) OLS computes the sample estimates which minimises the sum of squared deviation between predicted and actual values
- d) OLS computes the sample estimates which maximises the sum of squared deviation between predicted and actual values

8. Which of the following assumptions of multiple linear regression can be overlooked in data mining?

- a) Noise follows a normal distribution
- b) Observations are independent
- c) Linear relationship holds true
- d) Heteroscedasticity

9. Which of the following statements is incorrect with respect to explanatory modelling?

- a) The final model selected may not have the best predictive accuracy
- b) Full sample is used to estimate the final model
- c) There is no previously assumed structure
- d) None of the above

10 Which of the following visualisation techniques can be used for measuring predictive performance of a model?

- a) Decile chart
- b) Histogram
- c) Lift curve
- d) Both a and c

NPTEL Course: Business analytics and data mining Modeling using R

Week - 6, Assignment – 6

1. Which of the following is correct about the below given statements?

Assertion (S): The value of adjusted R^2 is always less than the value of R^2

Reason (R): Adjusted R^2 accounts for the number of predictors in multiple linear regression model.

- a) Both S and R are true and R is the correct explanation of S
- b) Both S and R are true but R is not the correct explanation of S
- c) S is true but R is false
- d) S is false but R is true

2. Which of the following is true about the best value of 'k' in kNN when working with data having complex and irregular structures?

- a) Value of 'k' should be on the higher side
- b) Value of 'k' should be on the lower side
- c) Value of 'k' should be equal to the total number of observations in the dataset
- d) The value of 'k' has no impact

3. Which of the following statements is incorrect with respect to adjusted R-squared value?

- a) Higher the number of predictors, higher the adjusted R-squared value
- b) Adjusted R-squared uses a penalty on the number of predictors
- c) Higher values of adjusted R-squares indicate better fit
- d) None of the above

4. Which of the following linear regression algorithms can be used for variable selection and dimension reduction?

- a) Exhaustive search
- b) Partial iterative search
- c) Both A and B
- d) None of the above

5. Which of the following partial iterative search algorithms start with the full model?

- a) Forward selection
- b) Backward selection
- c) Exhaustive search
- d) Stepwise regression

6. Which of the following algorithms overlooks the pairs or groups of predictors that perform well together but perform poorly as single predictors?

- a) Forward selection
- b) Backward selection
- c) Exhaustive search
- d) None of the above

7. What would be the Euclidean distance between the following data points with 4 predictors S (3,5,2,8) and T (1,4,6,2)
- a) 16.15
 - b) 7.54
 - c) 5
 - d) 13
8. Which of the following is highly likely when using a high value of k in k-NN technique?
- a) Fitting to local patterns
 - b) Fitting to global patterns
 - c) Fitting to noise
 - d) None of the above
9. Which of the following scenario is regarded as a naïve rule in k-NN?
- a) When $k = 1$
 - b) When $k > 1$
 - c) When $k = n$ (where 'n' is the number of total observations)
 - d) When $1 < k < n$ (where 'n' is the number of total observations)
10. Which of the following is true when k-NN is used for prediction tasks rather than classification tasks?
- a) Computation of distance between the new observation and training partition records is different
 - b) Value of new record is determined using weighted average of all the k-nearest records
 - c) Value of new record is determined using weighted average of the records belonging to the dominant class
 - d) Overall misclassification error is used as performance metric

NPTEL Course: Business analytics and data mining Modeling using R

Week-7, Assignment-7

1. Which technique has a step for finding training partition records that have the same predictor values as the new observation?
 - a) Naïve Bayes
 - b) k-NN
 - c) Complete Bayes
 - d) Multiple linear regression
2. Which of the following are correct about the statements given below?

Assertion (S): Complete Bayes model is not suitable for prediction task

Reason (R): It is difficult to find training records with exact values of predictors as in the new record

 - a) Both S and R are true and R is the correct explanation of S
 - b) Both S and R are true but R is not the correct explanation of S
 - c) S is true but R is false
 - d) S is false but R is true
3. A dataset consists of 800 records of whether a flight was delayed or not due to weather issues (clear weather or bad weather). Out of 800 flights, 600 flights were delayed and 200 flights were on time. Out of the total flights that got delayed, 450 flights faced bad weather. For the remaining flights which were on time, only 25 flights faced bad weather. Compute the conditional probability (exact Bayes) of the flights that got delayed given the weather was clear.
 - a) 0.25
 - b) 0.46
 - c) 0.75
 - d) 0.54
4. Which of the following statements related to Bayes model is incorrect?
 - a) Bayes model is more suitable for classification task rather than prediction task
 - b) When number of predictors is more, Exact Bayes model is more suitable
 - c) It is preferable to use categorical predictors for computing Bayes model
 - d) Numerical predictors should be preferably converted into categorical predictors for Bayes model
5. Which of the following are correct about the below given statements?

I: Naïve Bayes model is suitable for classification problems.

II: The probability values in Naïve Bayes, should be accurate in absolute terms.

 - a) Statement I is true and Statement II is false
 - b) Statement II is true and Statement I is false
 - c) Both the statements are true
 - d) Both the statements are false

6. Which of the following statements are true with respect to Naïve Bayes model?

- a) Predictor values must not be independent of each other for a given class
- b) Absolute accuracy in actual probability values is essential to classify a new record
- c) Reasonable accuracy in rank ordering of probability values is required to classify a new observation
- d) Computation of denominator in Naïve Bayes formula impacts the rank ordering of probability values

7. For which of the following scenarios, it is acceptable to misclassify a few truthful financial reports as fraudulent reports?

- a) When the goal is to reduce overall misclassification error
- b) When the goal is to accurately identify records belonging to a specific class of interest
- c) When the goal is to maximize overall classification accuracy
- d) None of the above

8. Which of the following about cut-off value is suitable for identifying maximum records belonging to a specific class of interest?

- a) Below 0.5
- b) Above 0.5
- c) Equal to 0.5
- d) None of the above

9. Which of the following tools are suitable for computing conditional probability?

- a) Pivot Table
- b) Conditional Formatting
- c) Pie chart
- d) None of the above

10. Which of the following are limitations of Naïve Bayes method?

- a) Performs well mainly for classification of records
- b) Estimation of actual probabilities of a class
- c) Requires large number of records to obtain good results
- d) None of the above

NPTEL Course: Business analytics and data mining Modeling using R
Week-8, Assignment-8

1. Which of the following partitions are utilized by pruning steps of decision trees?
 - a) Training partition
 - b) Validation partition
 - c) Test partition
 - d) Any part of the data

2. Which of the following statements is not true about recursive partitioning steps of decision trees?
 - a) Recursive partitioning results in non-overlapping multi-dimensional rectangles
 - b) Recursive partitioning results in smaller and smaller rectangular regions
 - c) Recursive partitioning continues till heterogeneous groups are reached
 - d) Recursive partitioning results in zero error

3. What will be the gini index value when there are two classes with equal representation of each class?
 - a) 0
 - b) 1
 - c) -1
 - d) 0.5

4. How many terminal nodes will be estimated in a decision tree with 'n' number of decision nodes?
 - a) n
 - b) (n-1)
 - c) (n+1)
 - d) Can't say

5. Which of the following R functions can be used to filter overlapping observations in training and validation partitions?
 - a) replace
 - b) intersect
 - c) rpart.control
 - d) split

6. Which of the following is used as the basis for finding a suitable split-value of a predictor?
- a) Impurity increase
 - b) Impurity reduction
 - c) size
 - d) None of the above
7. Which of the following data mining tasks can be modeled using decision trees?
- a) Classification task only
 - b) Prediction task only
 - c) Both classification and prediction task
 - d) None of the above
8. Which of the following steps is used to fit the decision tree to the predictors' information and not to the noise?
- a) Splitting
 - b) Pruning
 - c) Partitioning
 - d) None of the above
9. Compute the entropy measure value for the following case:
There are 12 students who had taken a test. But only 9 students could pass the test and the remaining failed.
- a) 0.727
 - b) 1.20
 - c) 0.189
 - d) 0.811
10. While growing decision trees, nodes are formed for each recursive partition within the multi-dimensional space of predictors. The final nodes of fully grown decision tree corresponding to the final homogenous groups are referred as:
- a) Decision nodes
 - b) Root node
 - c) Terminal nodes
 - d) None of the above

NPTEL Course: Business analytics and data mining Modeling using R

Week-9, Assignment-9

1. What is the range of gini values for the binary classification tasks?
 - a) $\{0, 0.5\}$
 - b) $\{0, (m-1)/m\}$
 - c) $\{0, 1\}$
 - d) $\{0, m/(m-1)\}$
2. What is the range of entropy values for the binary classification tasks?
 - a) $\{0, \log(m)\}$
 - b) $\{0, 0.5\}$
 - c) $\{0, 1\}$
 - d) None of the above.
3. What will be the entropy measure value when there are two classes with equal representation of each class?
 - a) 0
 - b) 1
 - c) -1
 - d) 0.5
4. Which of the following are true about both linear and logistic regression?
 - a) One or more independent variables
 - b) Same estimation method
 - c) Single dependent variable
 - d) None of the above
5. In binary classification, a cutoff value of 0.5 means that cases with an estimated probability, $P(Y=1) > 0.5$ are classified to:
 - a) Class 1
 - b) Class 0
 - c) Both a and b
 - d) None of the above
6. Which of the following are not true about logistic regression?
 - a) Least squares method is used
 - b) Maximum likelihood method is used
 - c) Instead of using outcome variable (y) in the model, a function of y, called logit is used
 - d) None of the above
7. Which of the following can be true when we add a new variable in the linear regression model?
 - a) R-squared and adjusted R-squared both increase
 - b) R-squared increases and adjusted R-squared decreases
 - c) R-squared decreases and adjusted R-squared decreases
 - d) R-squared decreases and adjusted R-squared increases

8. Which of the following is correct about the training of linear regression models based on below given statements?

I: Overfitting is more likely if we have less data

II: Overfitting is more likely when the set of all possible mappings of inputs to outputs is small.

- a) Both are False
- b) I is False and II is True
- c) I is True and II is False
- d) Both are True

9. Which type of data is typically modeled by regression trees?

- a) Linear
- b) Nonlinear
- c) Can't say
- d) None of the above

10. Which type of relationship between the input attribute and output attribute is assumed in simple regression?

- a) Linear
- b) Quadratic
- c) Inverse
- d) None of the above

NPTEL Course: Business analytics and data mining Modeling using R
Week-10, Assignment-10

1. Which of the following tasks are suitable for logistic regression?
 - a) Profiling
 - b) Classification
 - c) Both of them
 - d) None of the above
2. Which of the following estimation methods are used in logistic regression?
 - a) Ordinary least squares
 - b) Maximum likelihood
 - c) Partial least squares
 - d) None of the above
3. Which type of output values are predicted using logistic regression?
 - a) Continuous
 - b) Discrete
 - c) Range
 - d) None of the above
4. Which of the following are true about logistic regression?
 - a) $\text{Logit} = \text{Log}(\text{odds})$
 - b) $\text{Logit} = \text{Log}(P/(1-P))$
 - c) $P = \text{Logistic}(x)$
 - d) None of the above
5. Let us consider the following logistic regression model:
 $P(y=1 \mid x, w) = g(w_0 + w_1 \cdot x)$, where $g(z)$ is the logistic function.
What would be the range of P values in such a case?
 - a) $(-\infty, 0)$
 - b) $(0, 1)$
 - c) $(-\infty, \infty)$
 - d) $(0, \infty)$
6. What is the range of the logit function?
 - a) $(-\infty, 0)$
 - b) $(0, 1)$
 - c) $(-\infty, \infty)$
 - d) $(0, \infty)$
7. Which of the following is true about binary logistic regression?
 - a) Dependent variable is continuous.
 - b) Dependent variable has two classes.
 - c) There is no dependent variable.
 - d) None of the above.

8. Which of the following is considered a suitable goodness of fit metric for logistic regression?

- a) SSE
- b) Deviance
- c) Multiple R squared
- d) $1 - \text{Deviance} / \text{Null Deviance}$
- e) None of the above

9. Which of the following statements is true about logistic regression?

- a) Model parameters in the logit model are additive
- b) Model parameters in the logit model are multiplicative
- c) Model parameters in the odds model are additive
- d) Model parameters in the odds model are multiplicative

10. Which of the following correctly defines the logit?

- a) Logical operation
- b) Natural logarithm of odds ratio
- c) Natural logarithm of weighted inputs
- d) None of the above

NPTEL Course: Business analytics and data mining Modeling using R
Week-11, Assignment-11

1. What is the backpropagation algorithm?
 - 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 adjust the inputs
 - c) It is the transmission of error back through the network to allow weights to be adjusted so that the network can learn
 - d) None of the above
2. Which of the following defines neural networks?
 - a) Complex linear functions with many parameters
 - b) Complex nonlinear functions with many parameters
 - c) Complex discrete functions with many parameters
 - d) None of the above
3. Which of the following can be an application of neural network?
 - a) Sales forecasting
 - b) Fact checking
 - c) Risk analysis
 - d) None of the above
4. Which of the following is true about multilayer feedforward neural networks?
 - a) Not fully connected
 - b) Same transfer or activation function in each layer
 - c) Only one hidden layer is allowed
 - d) None of the above
5. Which of the following does not happen in the backpropagation algorithm?
 - a) Error in output is propagated backwards to update weight values
 - b) Error in output is propagated backwards to update bias values
 - c) Backpropagation is done iteratively for many epochs
 - d) None of the above
6. Which of the following is not true about multilayer feedforward neural networks?
 - a) Consists of three main types of layers
 - b) Input layer typically has input nodes equal to the number of input features
 - c) Hidden layers are allowed a fixed number of nodes only
 - d) None of the above
7. Which of the following is the most commonly used neural network architecture among feedforward networks?
 - a) Multi layer perceptron
 - b) Perceptron
 - c) Radial basis function network
 - d) None of the above

8. For which of the following neural networks, information does not move only in one direction?

- a) Multi layer perceptron
- b) Feedforward network
- c) Recurrent neural network
- d) None of the above

9. Which of the following is the recommended data normalization in neural networks?

- a) Z-score normalization or standardization
- b) Min-max normalization
- c) Both of them
- d) None of the above

10. What is the transfer function?

- a) It is used to add noise in the feedforward neural networks
- b) It is the transmission of error back through the network to adjust the inputs
- c) It is applied on the weighted sum of incoming values to produce the outgoing values for a layer
- d) None of the above

NPTEL Course: Business analytics and data mining Modeling using R
Week-12, Assignment-12

1. Which of the following data mining tasks should not be conducted using discriminant analysis?
 - a) Prediction
 - b) Classification
 - c) Clustering
 - d) None of the above
2. Which of the following is true about linear classification functions used in discriminant analysis?
 - a) Provide the basis for discrimination of records into classes
 - b) Linear functions of predictors that maximize ratio of between-class variability to within-class variability
 - c) Coefficients of linear discriminant are optimized w.r.t class separation
 - d) None of the above
3. Which of the following plot can be helpful in assessing class separation for discriminant analysis?
 - a) Histogram
 - b) Scatter plot
 - c) Bar chart
 - d) None of the above
4. What is the maximum number of needed discriminant functions when m classes are present?
 - a) m
 - b) $m-1$
 - c) $m/2$
 - d) None of the above
5. Which of the following is true assumption about correlation structure between predictors in discriminant analysis?
 - a) Different for each class
 - b) Same for each class
 - c) Does not matter
 - d) None of the above
6. Which of the following are true about discriminant analysis and linear regression?
 - a) Same estimation technique
 - b) Coefficients are optimized using same mechanism
 - c) Different estimation technique
 - d) None of the above
7. Which of the following updating mechanisms yields more accurate results in neural networks?
 - a) Batch updating
 - b) Case updating
 - c) Both a and b
 - d) None of the above

8. Which of the following is true about updating mechanisms in neural networks?
- a) Case updating is done after each case or record is run through the network.
 - b) Batch updating is done after each case or record is run through the network.
 - c) Batch updating is done after all records are run through the network.
 - d) None of the above
9. What is the basic advantage of data normalization step?
- a) Smaller values improve the model
 - b) Values falling in a smaller range improve the model
 - c) Computing performance is better
 - d) None of the above
10. Which stopping criteria are typically used in the training of neural networks?
- a) Small incremental change in bias and weight values
 - b) Rate of change of error function values reaches a required threshold
 - c) Limit on no. of runs is reached
 - d) None of the above