001. Which of the following is an advantage of stacking? **B**
  A   Less complexity   B   Better prediction
  C   Lower time of execution   D   Low Memory

002. SVM stands for? **B**
  A   Simple Vector Machine   B   Support Vector Machine
  C   Super Vector Machine   D   Small Vector Machine

003. SVM is a supervisedMachine Learningcan be used for _____ **C**
  A   Regression   B   Classification
  C   Either A or B   D   Clustering

004. In SVM, Linear separator,Hyper plane **C**
  A   f(x)=sign(w/x+b)   B   f(x)=sign(w+x+b)
  C   f(x)=sign(w*x+b)   D   f(x)=sign(w-x+b)

005. Which of the following is not a ensemble learning techniques? **D**
  A   Bagging   B   Boosting
  C   Stacking   D   PCA

006. Which of the following is true about averaging ensemble? **C**
  A   It can only be used in classification   B   It can only be used in regression
      problem   problem
  C   It can be used in both classification as   D   It cant be used in classification as
      well as regression   well as regression

007. Which of the following algorithm is not an example of an ensemble method? **D**
  A   Extra Tree Regressor   B   Random Forest
  C   Gradient Boosting   D   Decision Tree

008. Generally, an ensemble method works better, if the individual base models have **A**
  _____? Note: Suppose each individual base models have accuracy greater
  than 50%
  A   Less correlation among predictions   B   High correlation among predictions
  C   Correlation does not have any impact   D   None of the above
      on ensemble output

009. Suppose you are dealing with 4 class classification problem and you want to train a **D**
  SVM model on the data for that you are using One-vs-all method.How many times we
  need to train our SVM model in such case?
  A   1   B   2
  C   3   D   4

010. Suppose you have trained an SVM with linear decision boundary after training SVM, **C**
  you correctly infer that your SVM model is underfitting. Which of the following option
  would you more likely to consider iterating SVM next time?
  A   You want to increase your data points   B   You want to decrease your data
                                                    points
  C   You will try to calculate more   D   You will try to reduce the features
      variables

011. If a dataset can be classified into two classes by using a single straight line, is called **A**
  _____
  A   Linear SVM   B   Non- Linear SVM
  C   Kernel SVM   D   Cluster SVM

012. Nave Bayes is? **A**
  A   Conditional Independence   B   Conditional Dependence
  C   Un-conditional Independence   D   Un-conditional dependence

013. Support vector machines is _____ **C**
  A   Semi-supervised Machine Learning   B   Unsupervised Machine Learning
  C   Supervised Machine Learning   D   Reinforcement learning

014. The effectiveness of an SVM NOT depends upon **D**
  A   Selection of Kernel   B   Kernel Parameters
  C   Soft Margin Parameter C   D   Size of data set

015. In Hyper plane, f(x)=sign(w*x+b) where w is a? **B**
- A   Constant
- B   Vector
- C   Distance
- D   Width

016. Closest Point to the hyper plane are support vectors **A**
- A   True
- B   False
- C   Unpredictable
- D   True in few cases only

017. Which of the following is the consequence between a node and its predecessors while creating bayesian network? **A**
- A   Conditionally independent
- B   Functionally dependent
- C   Both Conditionally dependant&Dependant
- D   Dependent

018. Bayes rule can be used for _____ **C**
- A   Solving queries
- B   Increasing complexity
- C   Answering probabilistic query
- D   Decreasing complexity

019. Which of the following is NOT True about Ensemble Techniques? **D**
- A   Bagging decreases the variance of the classifier.
- B   Boosting helps to decrease the bias of the classifier.
- C   Bagging combines the predictions from different models and then finally gives the results.
- D   Bagging and Boosting are the only available ensemble techniques.

020. The formula for Bayes &#39 theorem is given as: **B**

$$P(A|B)= \frac{P(B|A)P(A)}{P(B)}$$

Where P(B|A) is _____?
- A   Posterior probability
- B   Likelihood probability
- C   Prior Probability
- D   Marginal Probability

021. In Naive Bayes, Numerical variable must be binned and converted to ? **A**
- A   Categorical Values
- B   Numerical Values
- C   Only zeros
- D   Only ones

022. _____ terms are required for building a bayes model. **C**
- A   1
- B   2
- C   3
- D   4

023. Naive Bayes requires? **B**
- A   Numerical Values
- B   Categorical Values
- C   Exponential Values
- D   in between 0 and 1 only

024. Spam Classification is an example for? **A**
- A   Naive Bayes
- B   Probabilistic condition
- C   Random Forest
- D   Logistic Regression

025. Which of the following is NOT true about Support Vector Regression? **D**
- A   It is a supervised learning algorithm that is used to predict discrete values
- B   It uses the same principle as the SVMs
- C   he basic idea behind SVR is to find the best fit line
- D   In SVR, the best fit line is the hyperplane that has the minimum number of points

026. Which of the following is not a machine learning algorithm? **A**
- A   SVG
- B   SVM
- C   Random forest
- D   SVR

027. Which of the following is NOT true about hyperplane in SVM? **C**
- A   This best boundary is known as the hyperplane of SVM.
- B   The dimensions of the hyperplane depend on the features present in the dataset
- C   If there are 2 features , then hyperplane will be a 2-dimension
- D   We always create a hyperplane that has a maximum margin

plane.

028. In SVM, The data points that are the closest to the hyperplane and which affect the position of the hyperplane are termed as _____ **B**
A   Soft margin   B   Support Vectors
C   Hard margin   D   Nearest Neighbors

029. What do you mean by generalization error in terms of the SVM? **B**
A   How far the hyperplane is from the support vectors   B   How accurately the SVM can predict outcomes for unseen data
C   The threshold amount of error in an SVM   D   How accurately the SVM can predict outcomes for trained data

030. The SVMs are less effective when _____ **C**
A   The data is linearly separable   B   The data is clean and ready to use
C   The data is noisy and contains overlapping points   D   The data has numerical values

031. Which of the following is the most popular choice for text classification problems? **D**
A   Decision tree   B   Logistic regression
C   Random forest   D   Nave bayes

032. Which of the following is a disadvantage of Nave bayes algorithm? **D**
A   It can be used for Binary as well as Multi-class Classifications   B   It performs well in Multi-class predictions as compared to the other Algorithms
C   It is the most popular choice for text classification problems   D   It assumes that all features are independent or unrelated, so it cannot learn the relationship between features.

033. Which of the followings in NOT an example of Nave Bayes Algorithm **D**
A   Spam filtration   B   Sentimental analysis
C   Classifying articles   D   House rent prediction

034. Nave Bayes algorithm is based on _____ and used for solving classification problems. **A**
A   Bayes Theorem   B   Candidate elimination algorithm
C   EM algorithm   D   S Algorithm

035. Which of the followings is not a type of Nave Bayes Model **D**
A   Gaussian   B   Multinomial
C   Bernoulli   D   Polynomial

036. How do we perform Bayesian classification when some features are missing? **C**
A   We assuming the missing values as the mean of all values.   B   We ignore the missing features.
C   We integrate the posteriors probabilities over the missing features.   D   Drop the features completely.

037. _____ provides way and means of weighing up the desirability of goals and the likelihood of achieving them. **A**
A   Utility theory   B   Decision theory
C   Bayesian networks   D   Probability theory

038. Probability provides a way of summarizing the _____ that comes from our laziness and ignorances. **B**
A   Belief   B   Uncertaintity
C   Joint probability distributions   D   Randomness

039. Time complexity for Naive Bayes classifier for n feature, L classdatais _____ **A**
A $n*L$ B $O(n+L)$
C $O(n*L)$ D $O(n/L)$

040. Naive Bayes pays attention to complex interactions and _____ **B**
A   Statistical Model   B   Local Structure

| | C | Dynamic Model | | D | Logical Structure |

041. Which of the following is true about weighted majority votes?1. We want to give higher **D** weights to better performing models 2. Inferior models can overrule the best model if collective weighted votes for inferior models is higher than best model 3. Voting is special case of weighted voting 4. We want to give low weights to better performing models

A	1 and 2	B 1 and 3
C	1, 2 and 4	D 1, 2 and 3

042. Which of the following are correct statement(s) about stacking? 1. A machine learning **C** model is trained on predictions of multiple machine learning models 2. A Logistic regression will definitely work better in the second stage as compared to other classification methods 3. First stage models are trained on full / partial feature space of training data 4. A machine learning model is trained on prediction of single machine learning model

A	1 and 2	B 2 and 3
C	1 and 3	D 2 and 4

043. Which of the following can be one of the steps in stacking?1. Divide the training data **A** into k folds 2. Train k models on each k-1 folds and get the out of fold predictions for remaining one fold 3. Divide the test data set in k folds and get individual fold predictions by different algorithms

A	1 and 2	B 2 and 3
C	1 and 3	D 1, 2 and 3

044. Disadvantages of Nave Bayes Classifier _____   **A**

A	Naive Bayes assumes that all features are independent or unrelated, so it cannot learn the relationship between features.	B	It performs well in Multi-class predictions as compared to the other Algorithms.
C	Nave Bayes is one of the fast and easy ML algorithms to predict a class of datasets.	D	It is the most popular choice for text classification problems.

045. Suppose you are given n predictions on test data by n different models (M1, M2, . Mn) **D** respectively. Which of the following method(s) can be used to combine the predictions of these models? Note: We are working on a regression problem 1. Median 2. Product 3. Average 4. Mode

A	1, 2	B 2, 3 and 4
C	3, 4	D 1, 2 and 3

046. How can we assign the weights to output of different models in an ensemble?1. Use an **C** algorithm to return the optimal weights 2. Choose the weights using cross validation 3. Give low weights to more accurate models 4. Give high weights to more accurate models

A	1 and 2	B 1 and 3
C	1, 2 and 4	D 1, 2 and 3

047. What is true about an ensembled classifier?1. Classifiers that are more sure can vote **D** with more conviction 2. Classifiers can be moresureabout a particular part of the space 3. Most of the times, it performs better than a single classifier

A	1 and 2	B 1 and 3
C	2 and 3	D 1, 2 and 3

048. Which of the following option is / are correct regarding benefits of ensemble model?1. **C** Better performance 2. Generalized models 3. Better interpretability

A	1 and 3	B 2 and 3
C	1 and 2	D 1, 2 and 3

049. Which of the following method is used for finding optimal of cluster in K-Mean **A** algorithm?

A	Elbow method	B Manhattan method

C	Ecludian method	D	Gaussian method	

050. _____is basically a type of unsupervised learning method **B**
A	Unsupervised learning	B	clustering
C	semi supervised	D	classification

051. Feature scaling is an important step before applying K-Mean algorithm. What is reason **A**
behind this?
A	In distance calculation it will give the same weights for all features	B	You always get the same clusters. If you use or dont use feature scaling
C	In Manhattan distance it is an important step but in Euclidian it is not	D	For easy remembering the values

052. In the Naive Bayes algorithm, suppose that prior for class w1 is greater than class w2, **B**
would the decision boundary shift towards the region R1(region for deciding w1) or
towards region R2(region for deciding w2)?
A	towards region R1.	B	towards region R2.
C	No shift in decision boundary.	D	It depends on the exact value of priors.

053. Which of the following statement is TRUE about the Bayes classifier? **A**
A	Bayes classifier works on the Bayes theorem of probability.	B	Bayes classifier is an unsupervised learning algorithm.
C	Bayes classifier is also known as maximum apriori classifier.	D	It assumes the independence between the independent variables or features.

054. What is the minimum no. of variables/ features required to perform clustering? **B**
A	0	B	1
C	2	D	3

055. Which of the following is/are valid iterative strategy for treating missing values before **C**
clustering analysis?
A	Imputation with mean	B	Nearest Neighbor assignment
C	Imputation with Expectation Maximization algorithm	D	Imputation with mode

056. Which of the following techniques would perform better for reducing dimensions of a **A**
data set?
A	Removing columns which have too many missing values	B	Removing columns which have high variance in data
C	Removing columns with dissimilar data trends	D	Removing columns which have too many NULL values

057. Which of the following algorithms cannot be used for reducing the dimensionality of **D**
data?
A	t-SNE	B	PCA
C	LDA False	D	PAC

058. Which of the following are the high and low bounds for the existence of F-Score? **A**
A	[0,1]	B	(0,1)
C	[-1,1]	D	(-1,0)

059. _____is the task of dividing the population or data points into a number of groups. **B**
A	Unsupervised learning	B	clustering
C	semi supervised	D	classification

060. Which of the following is not clustering method? **D**
A	Density-Based	B	Hierarchical Based
C	Grid-based	D	Project Based

061. _____consider the clusters as the dense region having some similarity and **A**
different from the lower dense region of the space
A	Density-Based	B	Hierarchical Based
C	Grid-based	D	Distance based

062. Agglomerative has _____ approach **B**

	A	top down	B	bottom up
	C	dynamic	D	static

063. Principal Component Analysis is a/an _____ learning algorithm that is used for the dimensionality reduction inmachine learning   **A**

A	Unsupervised	B	Supervised
C	Semi-supervised	D	Reinforcement

064. PCA works better if there is?A linear structure in thedataIf the data lies on acurved surface and not on a flat surfaceIf variables are scaled inthe same unit   **C**

A	1 and 2	B	2 and 3
C	1 and 3	D	1 ,2 and 3

065. What is the importance of using PCA before the clustering?   **C**

A	Avoid bad features	B	Find the explained variance
C	Find which dimension of data maximize the features variance	D	Find good features to improve your clustering score

066. When uses PCA?   **A**

A	You want to find latent features and reduce dimensionality	B	When my data is small and with a few features
C	When I have a overfit case	D	Everytime before uses a Machine Learning algorithm

067. Why you have to drop unimportant features?   **B**

A	Find the correct clusters	B	Using the most important features will give you better efficiency predicting your target
C	Standardize the data	D	To trains the model faster

068. What PCA does afterfall?   **A**

A	Reduce dimensionality of the data and create new features from features set given	B	Create clusters in order to let you know what are the class
C	Predicts your target with high efficiency	D	Give you the highest number of features possible, to maximize the efficience of your Machine Learning algorithm

069. Following the steps to run a PCA &#39s algorithm, why is so important standardize your data?   **C**

A	Standardize data allows other people understand better your work	B	Make the training time more fast
C	Find the features which can best predicts Y	D	Use the best practices of data wrangling

070. The PCA algorithm is based on some mathematical concepts such as _____   **D**

A	Variance	B	Covariance
C	Eigen values and Eigen vectors	D	Confusion matrix

071. A matrix containing the covariance between the pair of variables is called _____   **A**

A	Covariance Matrix	B	Confusion matrix
C	Correlation matrix	D	Error matrix

072. Which of the following are not an applications of PCA?   **D**

A	image processing	B	movie recommendation system
C	optimizing the power allocation in various communication channels	D	Relation between height and weight

073. Which of the following options are correct, when you are applying PCA on a image dataset?1. It can be used to effectively detect deformable objects. 2. It is invariant to affine transforms. 3. It can be used for lossy image compression. 4. It is not invariant to shadows.   **C**

A	1 and 2	B	2 and 3
C	3 and 4	D	1 and 4

074. Which of the following option is true?   **A**
- A   LDA explicitly attempts to model the   B   Both attempt to model the difference
      difference between the classes of   between the classes of data.
      data. PCA on the other hand does not
      take into account any difference in
      class.
- C   PCA explicitly attempts to model the   D   Both dont attempt to model the
      difference between the classes of   difference between the classes of
      data. LDA on the other hand does not   data.
      take into account any difference in
      class.

075. Imagine you are dealing with 10 class classification problem and you want to know that **B**
at most how many discriminant vectors can be produced by LDA. What is the correct
answer?
- A   20   B   9
- C   21   D   11

076. PCA generally tries to find the _____ dimensional surface to project the _____   **A**
dimensional data.
- A   Lower, higher   B   Higher, lower
- C   Lower, lower   D   Upper, upper

077. Which of the following statement is true for a t-SNE cost function?   **B**
- A   It is asymmetric in nature.   B   It is symmetric in nature.
- C   It is same as the cost function for   D   It is same as t-test
      SNE.

078. Which of the following statement is correct for t-SNE and PCA?   **D**
- A   t-SNE is linear whereas PCA is non-   B   t-SNE and PCA both are linear
      linear
- C   t-SNE and PCA both are nonlinear   D   t-SNE is nonlinear whereas PCA is
                                                     linear

079. Which of the following is/are not true about Centroid based K-Means clustering   **D**
algorithm and Distribution based expectation-maximization clustering algorithm
- A   Both starts with random initializations B   Both are iterative algorithms
- C   Both have strong assumptions that   D   Expectation maximization algorithm is
      the data points must fulfill   a special case of K-Means

080. Which of the following algorithm is most sensitive to outliers?   **A**
- A   K-means clustering algorithm   B   K-medians clustering algorithm
- C   K-modes clustering algorithm   D   K-centered clustering algorithm

081. In the figure below, if you draw a horizontal line on y-axis for y=2. What will be the   **B**
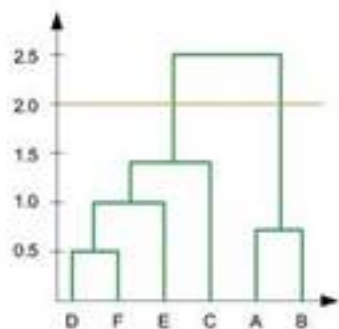

number of clusters formed?
- A   1   B   2
- C   3   D   4

082. If you are using Multinomial mixture models with the expectation-maximization   **C**
algorithm for clustering a set of data points into two clusters, which of the assumptions
are important:

| A | All the data points follow two Gaussian distribution | B | All the data points follow n Gaussian distribution (n >2) |
| C | All the data points follow two multinomial distribution | D | All the data points follow n multinomial distribution (n >2) |

083. The most popularly used dimensionality reduction algorithm is Principal Component **C**
Analysis (PCA). Which of the following is/are true about PCA?PCA is anunsupervised
methodIt searches forthe directions that data have the largest varianceMaximum
numberof principal components <= number of featuresAll principalcomponents are not
orthogonal to each other

| A | 1,2 | B | 2,3 |
| C | 1,2,3 | D | 1,2,4 |

084. After performing K-Means Clustering analysis on a dataset, you observed the following **D**
dendrogram. Which of the following conclusion can be drawn from the
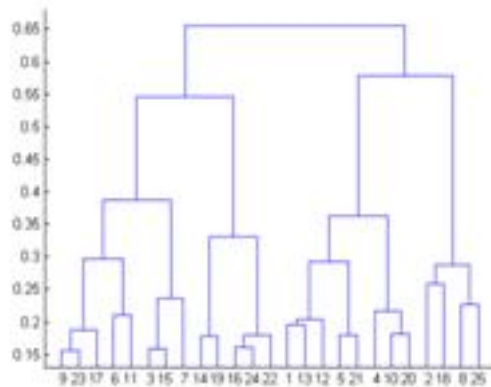
dendrogram?

| A | There were 28 data points in clustering analysis | B | The best no. of clusters for the analyzed data points is 4 |
| C | The proximity function used is Average-link clustering | D | The above dendrogram interpretation is not possible for K-Means clustering analysis |

085. How can Clustering (Unsupervised Learning) be used to improve the accuracy of **C**
Linear Regression model (Supervised Learning)Creating different modelsfor different
cluster groups.Creating an input featurefor cluster ids as an ordinal variable.Creating
an input featurefor cluster centroids as a continuous variable.Creating an input
featurefor cluster size as a continuous variable.

| A | 1 only | B | 1 and 2 |
| C | 1, 2, 3 and 4 | D | 1, 2 and 3 |

086. Which of the following clustering algorithms suffers from the problem of convergence at **D**
local optima?K- Means clusteringalgorithmAgglomerative
clusteringalgorithmExpectation-Maximizationclustering algorithmDiverse
clusteringalgorithm

| A | 1 only | B | 2 and 3 |
| C | 2 and 4 | D | 1 and 3 |

087. In which of the following case LDA will fail? **A**

| A | If the discriminatory information is not in the mean but in the variance of the data | B | If the discriminatory information is in the mean but not in the variance of the data |
| C | If the discriminatory information is in the mean and variance of the data | D | If the discriminatory information is not in the mean and not in thevariance of the data |

088. Which of the following comparison(s) are true about PCA and LDA?Both LDA and PCA **D**
arelinear transformation techniquesLDA is supervised whereasPCA is
unsupervisedPCA maximize the varianceof the data, whereas LDA maximize the
separation between differentclasses

A	1 and 2	B	2 and 3
C	1 and 3	D	1,2 and 3

089. What will happen when eigenvalues are roughly equal?   **B**

A	PCA will perform outstandingly	B	PCA will perform badly
C	PCA will have average performance	D	Cant Say

090. Which of the following can act as possible termination conditions in K-Means?1.For a   **C**
fixed number of iterations. 2.Assignment of observations to clusters does not change
between iterations. Except for cases with a bad local minimum. 3.Centroids do not
change between successive iterations. 4.Terminate when RSS falls above a threshold.

A	1,2	B	2,3
C	1,2,3	D	1,3,4

091. Which of the following is/are not true about DBSCAN clustering algorithmFor data   **D**
points to be in acluster, they must be in a distance threshold to a core pointIt has
strong assumptionsfor the distribution of data points indataspaceIt has substantially
hightime complexity of order $O(n^3)$It does not require priorknowledge of the no. of
desired clusters

A	1 only	B	2 only
C	4 only	D	2 and 3

092. What is the trend in software nowadays?   **A**

A	to bring computer more &more closer to user	B	to solve complex problems
C	to be task specific	D	to be versatile

093. In which of the following scenarios is t-SNE better to use than PCA for dimensionality   **C**
reduction while working on a local machine with minimal computational power?

A	Dataset with 1 Million entries and 300 features	B	Dataset with 100000 entries and 310 features
C	Dataset with 10,000 entries and 8 features	D	Dataset with 10,000 entries and 200 features

094. What could be the possible reason(s) for producing two different dendrograms using   **B**
agglomerative clustering algorithm for the same dataset?1. Proximity function used 2.
of data points used 3. of variables used

A	Only 1	B	1, 2 and 3
C	Only 2	D	1 and 2

095. In which of the following cases will K-Means clustering fail to give good results? Data   **A**
points with outliersData points with differentdensitiesData points with roundshapesData
points with non-convexshapes

A	1, 2 and 4	B	2 and 3
C	2 and 4	D	1 and 2

096. Which of the following metrics, do we have for finding dissimilarity between two   **D**
clusters in hierarchical clustering?Single-linkComplete-linkAverage-link

A	1 and 2	B	1 and 3
C	2 and 3	D	1, 2 and 3

097. Which of the following sequences is correct for a K-Means algorithm using Forgy   **A**
method of initialization? Specify the number ofclustersAssign cluster
centroidsrandomlyAssign each data point tothe nearest cluster centroidRe-assign each
point tonearest cluster centroidsRe-compute clustercentroids

A	1, 2, 3, 5, 4	B	1, 3, 2, 4, 5
C	2, 1, 3, 4, 5	D	3, 4, 5, 2, 1

098. What are dendrites?   **A**

A	fibers of nerves	B	nuclear projections
C	other name for nucleus	D	error in ANN

099. What is shape of dendrites like   **C**

A	oval	B	round
C	tree	D	rectangular

100. The fundamental unit of network is — **C**
A brain
B nucleus
C neuron
D axon

101. Whats the main point of difference between human & machine intelligence? — **A**
A human perceive everything as a pattern while machine perceive it merely as data
B human have emotions
C human have more IQ & intellect
D human have sense organs

102. What is auto-association task in neural networks? — **B**
A find relation between 2 consecutive inputs
B related to storage & recall task
C predicting the future inputs
D reducing features

103. Example of a unsupervised feature map? — **B**
A text recognition
B voice recognition
C image recognition
D image classification

104. What is plasticity in neural networks? — **A**
A input pattern keeps on changing
B input pattern has become static
C output pattern keeps on changing
D output is static

105. John hopfield was credited for what important aspec of neuron? — **C**
A learning algorithms
B adaptive signal processing
C energy analysis
D searching algorithms

106. What is ART in neural networks? — **C**
A automatic resonance theory
B artificial resonance theory
C adaptive resonance theory
D abstract resonance theory

107. Who proposed the first perceptron model in 1958? — **D**
A McCulloch-pitts
B Marvin Minsky
C Hopfield
D Rosenblatt

108. Where does the chemical reactions take place in neuron? — **C**
A dendrites
B axon
C synapses
D nucleus

109. What is purpose of Axon? — **C**
A receptors
B transmitter
C transmission
D generator

110. Operations in the neural networks can perform what kind of operations? — **D**
A serial
B parallel
C random
D serial or parallel

111. Which action is faster pattern classification or adjustment of weights in neural nets? — **A**
A pattern classification
B adjustment of weights
C equal
D either of them can be fast, depending on conditions

112. Which of the following is NOT an example of deep learning? — **B**
A Self-driving cars
B Home price prediction
C Natural language processing
D Pattern recognition

113. Autoencoder is an example of _____ — **A**
A Deep learning
B Machine learning
C Data mining
D Data warehousing

114. The procedure to incrementally update each of weights in neural is referred to as? — **D**
A synchronization
B learning law
C learning algorithm
D both learning algorithm & law

115. What is an activation value? — **A**
A weighted sum of inputs
B threshold value
C main input to neuron
D initial value of neuron

116. Positive sign of weight indicates? — **A**
A excitatory input
B inhibitory input

C	can be either excitatory or inhibitory as such	D	excitatory output	

117. Negative sign of weight indicates? **B**

A	excitatory input	B	inhibitory input
C	excitatory output	D	inhibitory output

118. The amount of output of one unit received by another unit depends on what? **D**

A	output unit	B	input unit
C	activation value	D	weight

119. What is gradient descent? **B**

A	Activation function	B	Optimization algorithm
C	Loss function	D	Cost function

120. What does a gradient descent algorithm do? **C**

A	Tries to find the parameters of a model that minimizes the cost function	B	Adjusts the weights at the input layers
C	Both A and B	D	Neither A nor B

121. Which of the following loss function is used in regression? **C**

A	Logarithmic loss	B	Cross entropy
C	Mean squared error	D	sigmoid function

122. Which of the following deep learning models uses back propagation? **C**

A	Convolutional Neural Network	B	Multilayer Perceptron Network
C	Recurrent Neural Network	D	Reinforcement Network

123. Which of the following is not a method used to prevent overfitting in a neural network? **D**

A	Dropout of neurons	B	Early stopping
C	Batch normalization	D	one hot encoding

124. Which of the following is NOT a deep learning library? **B**

A	Tensorflow	B	Pandas
C	PyTorch	D	Keras

125. Which of the following is true about bias? **C**

A	Bias is inherent in any predictive model	B	Bias impacts the output of the neurons
C	Both A and B	D	Neither A nor B

126. In TensorFlow, what is the use of asession? **B**

A	The current work space session for storingthe code	B	We launch the graph in a session
C	A session is used to download the data	D	A session is used for exporting data out ofTensorFlow

127. How can we improve the calculation speed in TensorFlow, without losing accuracy? **A**

A	Using GPU	B	By doing random sampling on Tensors
C	By removing few nodes from computationalgraphs	D	By removing the hidden layers

128. How calculations work in TensorFlow **C**

A	Through vector multiplications	B	Through RDDs
C	Through Computational Graphs	D	Through map reduce tasks

129. Which of the following activation function can not be used in the output layer of an image classification model? **A**

A	ReLu	B	Softmax
C	Sigmoid	D	Log function

130. Suppose you have a dataset from where you have to predict three classes. Then which of the following configuration you should use in the output layer? **A**

A	Activation function = softmax, loss function = cross entropy	B	Activation function = sigmoid, loss function = cross entropy
C	Activation function = softmax, loss	D	Activation function = sigmoid, loss

function = mean squared error   function = mean squared error

131. Which tool is best suited for solvingDeep Learning problems?   **D**
  A   R   B   Sklearn
  C   Excel   D   TensorFlow

132. A tensor is similar to   **A**
  A   Data Array   B   ANN Model
  C   SQL query   D   Pythoncode

133. How do you feed external data intoplaceholders?   **B**
  A   by using import data command   B   by using feed_dict
  C   by using read data function   D   Not possible

134. Can we use GPU for faster computations in TensorFlow   **D**
  A   No, not possible   B   Possible only on cloud
  C   Possible only with small datasets   D   Yes, possible

135. Why Tensorflow uses computation algraphs?   **D**
  A   Tensors are nothing but computationalgraphs   B   Gaphs are easy to plot
  C   There is no such concept of computationalgraphs in TensorFlow   D   Calculations can be done in parallel

136. Kears is a deep learning framework on which tool   **B**
  A   R   B   TensorFlow
  C   SAS   D   Azure

137. What is the meaning of model=sequentil() in Keras?   **C**
  A   No such code in Keras   B   Keras should be used only for sequentialmodels like RNNs
  C   Keras builds sequential models   D   Creates a computational graph

138. Which tool is NOT Suited for buildingANN models   **C**
  A   Python   B   TensorFlow
  C   Excel   D   Keras

139. Can we have multidimensional tensors   **C**
  A   No tensor can have maximum two dimensions   B   Possible only in image data
  C   Yes possible   D   Possible only in geo tagged data

140. TensorFlow supports which python version?   **D**
  A   Python 3.0   B   Python 3.3
  C   Python 3.5   D   Python 3.63.9

141. There are .... main tensor type you can create in TensorFlow.   **C**
  A   2   B   3
  C   4   D   5

142. TensorFlow was initially released in ....   **A**
  A   November 9, 2015   B   November 8, 2015
  C   October 9, 2015   D   November 9, 2016

143. Which tool is a deep learning wrapper on TensorFlow   **B**
  A   Python   B   Keras
  C   PyTourch   D   Azure

144. How deep learning models are built on Keras   **A**
  A   by using sequential models   B   by using feed_dict
  C   by creating place holders and computationalgraphs   D   by creating data frames

145. TensorFlow is a free and open-source .... based library for machine learning.   **A**
  A   Python   B   Java
  C   PHP   D   Angular

146. TensorFlow is developed by ....   **C**
  A   IBM Team   B   Microsoft Team
  C   Google Brain team   D   None of the above