001.	Back-propagation networks are _______ Neural networks				**C**
	A	Feed back	B	Feed forward	
	C	Multilayer Feed forward	D	Single layer Feed back	
002.	During the forward pass of back-propagation learning, ________ is applied to the sensory nodes of the network.				**A**
	A	Input Vector	B	Hidden Vector	
	C	Output Vector	D	Error Vector	
003.	During the backward pass of back-propagation learning, ________ is propagated through the network.				**D**
	A	Input	B	Hidden Vector	
	C	Output	D	Error	
004.	The perceptron learning algorithm work with ___				**A**
	A	threshold Function	B	linear neurons	
	C	non-linear neurons	D	predicate logic	
005.	Multi layer feedforward networks are trained by using ______.				**C**
	A	Hebbian learning rule	B	Perceptron learning Rule	
	C	Error Correlation Learning Rule	D	Delta learning rule	
006.	Back propagation algorithm may be viewed as a generalization of ___ algorithm.				**D**
	A	Hebb	B	Widrow-Hoff	
	C	Delta	D	LMS	
007.	Backpropagation algorithm is _______ in nature.				**C**
	A	Definitive	B	Certain	
	C	Stochastic	D	Obvious	
008.	In, Backpropagation algorithm ____________ is to be selected carefully to ensure that the weights quickly converge to a response, without oscillations.				**A**
	A	Learning rate	B	Initial weights	
	C	Bias	D	Number of Input units	
009.	Multi layer perceptron is trained by using ______ Learning				**A**
	A	Supervised	B	Unsupervised	
	C	Semi-supervised	D	Reinforced	
010.	Which of the following properties is considered as the corner stone of neural network theory?				**A**
	A	Universal function approximation	B	Universal output approximation	
	C	Error minimization	D	Input generalization	
011.	What is the objective of backpropagation algorithm?				**C**
	A	to develop learning algorithm for multilayer feedforward neural network	B	to develop learning algorithm for single layer feedforward neural network	
	C	to develop learning algorithm for multilayer feedforward neural network, so that network can be trained to capture the mapping implicitly	D	none of the mentioned	
012.	What is true regarding backpropagation rule?				**B**
	A	it is a feedback neural network	B	actual output is determined by computing the outputs of units for each hidden layer	
	C	hidden layers output is not all important, they are only meant for supporting input and output layers	D	it is based on widrow-hoff learning rule	
013.	What is meant by generalized in statement backpropagation is a generalized delta rule ?				**A**
	A	because delta rule can be extended to hidden layer units	B	because delta is applied to only input and output layers, thus making it	

			more simple and generalized	
C	it has no significance	D	none of the mentioned	

014. The property that states that, a network having adequate number of neurons can approximate the underlying function to any arbitrary degree of accuracy is __ **A**

A Universal function approximation — B Universal output approximation
C Error minimization — D Input generalization

015. ___ is used for increasing the rate of learning while maintaining stability during weight updation. **C**

A Input size — B Learning rate
C Momentum — D Bias

016. Computational capability of a single layer perceptron is limited because of the__ **A**

A Nature of the Activation function — B Nature of the input
C Nature of the output — D Choice of initial weights

017. The LMS learning algorithm work with ___ **B**

A threshold logic — B linear neurons
C non-linear neurons — D predicate logic

018. Computational capability of a single layer perceptron is limited because of the__ **A**

A Lack of layered architecture — B Nature of the input/output
C Nature of the Bias — D Choice of initial weights

019. Which of the following tasks cannot be performed with backpropagation algorithm? **D**

A pattern mapping — B function approximation
C prediction — D pattern storage

020. Backpropagation algorithm uses __ to change the weights in the network **A**

A error gradient for a pattern — B some of input vectors
C target output — D bias

021. Does backpropagation learning is based on gradient descent along error surface? **A**

A Yes — B No
C cannot be said — D it depends on gradient descent but not error surface

022. The backpropagation law is also known as generalized delta rule, is it true? **A**

A Yes — B no
C cannot be said — D it depends on input

023. The backpropagation law is also known as ___ rule **A**

A generalized delta — B extended Hebbian law
C Hebbian law — D Perceptron law

024. Which of the following is not a limitation of back propagation rule? **D**

A local minima problem — B slow convergence
C scaling — D Poor computational capability

025. Does back propagation learning is based on gradient descent along error surface? **A**

A Yes — B No
C cannot be said — D it depends on gradient descent but not error surface

026. Multilayer Perceptron utilizes ______ learning technique. **A**

A Supervised — B Unsupervised
C Semi-supervised — D Reinforced

027. Introducing a bipolar signal function such as the hyperbolic tangent function can cause a significant ________. **C**

A Reduction of storage space — B Stability in the weight updation
C speed up in the network convergence — D Increase in the accuracy of prediction

028. In a backpropagation network, hidden layer neurons are ___ **B**

A Linear — B Sigmoidal
C Circular — D Elliptical

029. The error signal at the output of neuron j at iteration n is defined by __________. Where $d_j(n)$ refers to the desired response and $y_j(n)$ refers to the function signal **B**

appearing at the output of neuron n.

A $e_j(n) = y_j(n) - d_j(n)$ B $e_j(n) = d_j(n) - y_j(n)$
C $e_j(n) = d_j(n) + y_j(n)$ D $e_j(n) = d_j(n) * y_j(n)$

030. In sequential / pattern mode, weight updation is performed after the presentation of **A**
A each training example B all training examples
C few training examples of an epoch D first training example

031. In batch mode, weight updation is performed after the presentation of **B**
A each training example B all training examples of an epoch
C few training examples of an epoch D first training example

032. If average gradient value fall below a preset threshold value, the process may be __. **A**
A Terminated B Corrupted
C Extended infinitely D unresolvable

033. In order to maintain a smooth trajectory in weight space, in the back propagation algorithm, ___ has to be kept small. **B**
A Input size B Learning rate
C Momentum D Bias

034. How can learning process be stopped in backpropagation rule? **C**
A there is convergence involved B no heuristic criteria exist
C on basis of average gradient value D none of the mentioned

035. In a backpropagation network, input layer neurons are ___ **A**
A Linear B Sigmoidal
C Circular D Non linear

036. Learning rate in the back propagation algorithm with pattern update has to be kept __in order to maintain a smooth trajectory in weight space. **A**
A Small B Large
C Average D Constant

037. If the gradient has different signs on consecutive iterations then the momentum causes ___ in the weight space. **B**
A Acceleration B Deceleration
C Either Acceleration or deceleration D Cannot be determined

038. When neurons saturate the signal values are close to the extremes __ **A**
A 0 or 1 B +ve or -ve
C Cannot be determined D Infinite values

039. Incorrect choice of weights can lead to ___ **D**
A Faster convergence B Slower convergence
C Network paralysis D Network saturation

040. ____ of weights can avoid incorrect saturation problems. **D**
A Standardization B Normalization
C Maximization D Randomization

041. Large values of momentum cause the training algorithm to find __ **B**
A Deeper global minima B Deeper local minima
C Shallow local minima D Shallow global minima

042. Initialization of the weights of the entire network to the same value can lead to ___ **C**
A Faster convergence B Slower convergence
C Network paralysis D Network saturation

043. When the training data are redundant, ________ mode is advantageous. **A**
A Sequential B Batch
C Either sequential or batch D Neither sequential nor batch

044. During weight updation, the magnitude of the momentum should be __ if the weights are expected to have convergent dynamics. **A**
A Less than 1 B Greater than 1
C Equal to 1 D Greater than or equal to 1

045. If the gradient has same sign on consecutive iterations then the momentum causes ___ in the weight space. **A**

A Acceleration | B Deceleration
C Either Acceleration or deceleration | D Cannot be determined

046. Feedback connection strength are usually ? **A**

A Fixed | B Variable
C both fixed or variable type | D set to either 1 or 0

047. In a backpropagation network, output layer neurons are ___ **B**

A Linear | B Sigmoidal
C Circular | D Non linear

048. Use of hyperbolic tangent function comes with the advantage that the range of valid desired signals extends to **A**

A [1 + , 1] | B [, n]
C [n + , n] | D [1 + ,]

049. Feedforward networks are used for? **D**

A Autoassociation | B pattern storage
C both autoassociation& pattern storage | D pattern classification

050. _______ theorem determines the minimum number of hidden layers in a multilayer perceptron with an input-output mapping that provides an approximate realization of any continuous mapping. **A**

A Universal approximation | B Uniform approximation
C Function approximation | D Continuous approximation

051. Feedforward networks are not used for? **D**

A pattern mapping | B pattern association
C pattern classification | D Pattern Storage

052. Gradient of the pattern error is employed in weight updation in order to reduce the ______ over the entire training set. **C**

A Local Error | B Local minima
C Global Error | D Global minima

053. Which of the following is not a criterion for terminating network training? **D**

A Compare absolute value of squared error averaged over one epoch, with a training tolerance | B use the absolute rate of change of the mean squared error per epoch
C Check if Euclidean norm of the error gradient falls below a sufficiently small threshold | D Compare the number of patterns fed into the network with a threshold

054. The network generalizes well if it is able to **B**

A Predict incorrect or near incorrect outputs for unseen inputs | B Predict correct or near correct outputs for unseen inputs
C Complete the training in a considerably few iterations | D Maximize the error gradient

055. In a network with two hidden layers, the second hidden layer extracts **D**

A Local features of the function | B Features of Input Vector
C Features of Output Vector | D Global features of the function

056. RBF network consists of______ layers. **C**

A Single | B Finite number
C Three | D Multiple

057. In a RBF network, the hidden nodes implement _____ functions **D**

A Square | B Sigmoid
C Logistic | D Gaussian

058. In a RBF network, the output nodes implement _____ functions. **A**

A Linear Summation | B Sigmoid
C Logistic | D Radial Basis

059. Curve-fitting process is easier with ______. **B**

A One hidden layer | B Two hidden layers

C	Three hidden layers	D	N hidden layers	

060. In a network with two hidden layers, the first hidden layer extracts **A**

A	Local features of the function	B	Features of Input Vector
C	Features of Output Vector	D	Global features of the function

061. Network saturation is a condition where ____. **A**

A	Weight changes are almost negligible over consecutive epochs.	B	Weights are set to maximum values
C	Weights are set to minimum values	D	Weights are set to fixed values

062. An optimal learning rate reaches the error minimum in **B**

A	2 learning steps	B	single learning step
C	Finite number of learning steps	D	Infinite number of learning steps

063. Learning rates that are larger than twice the optimal value will **A**

A	Diverge from the solution	B	Converge to the error minimum
C	Converge to the global optimum	D	Take longer to converge

064. In RBF network, output layer is ______. **A**

A	Linear	B	Non-linear
C	Sigmoidal	D	Either sigmoidal or linear

065. RBF network contains _______ hidden layer(s) **A**

A	1	B	2
C	3	D	N

066. RBF networks are used for **C**

A	Pattern Association	B	Pattern Storage
C	Pattern Classification	D	Pattern clustering

067. Which of the following Radial Basis Functions is not covered by Micchellis theorem? **C**

A	Multiquadrics	B	Inverse multiquadrics
C	Inverse Quadratic	D	Gaussian

068. _______ are very good at interpolation. **C**

A	Single layer Perceptrons	B	Multilayer Perceptrons
C	RBF Networks	D	Feedback Networks

069. In RBF network, hidden layer is ______. **B**

A	Linear	B	Non-linear
C	Sigmoidal	D	Either sigmoidal or linear

070. The RBF network training is divided into _____stages **A**

A	Two	B	Three
C	Four	D	N(A finite number)

071. In the first stage of RBF network training, the weights from the ______layer are determined. **A**

A	Input to hidden	B	Hidden to output
C	Input to output	D	First hidden to second hidden

072. In the second stage of RBF network training, the weights from the ______layer are determined **B**

A	Input to hidden	B	Hidden to output
C	Input to output	D	First hidden to second hidden

073. A radial basis function network requires __ number of radial basis functions for N data points. **D**

A	1	B	2
C	3	D	N

074. The problem of reconstructing the hypersurface is said to be well posed even if the following condition is not satisfied. **C**

A	Existence	B	Uniqueness
C	Linearity	D	Continuity

075. The physical phenomenon responsible for generating the training data is a ___ problem. **A**

A	Well posed direct	B	Well posed inverse

C Ill posed direct | D Ill posed inverse

076. In a radial basis function network, number of radial basis functions is ___ number of data points **C**

A Less than
B Greater than
C Equal to
D Not dependent

077. For which of the following functions, the interpolation matrix is positive definite? **B**

A Multiquadrics
B Inverse multiquadrics
C Inverse Quadratic
D Either Inverse Quadratic or Multiquadrics

078. Which of the following functions is suitable for use in the design of RBF networks? **A**

A Multiquadrics
B Inverse multiquadrics
C Inverse Quadratic
D Either Inverse Quadratic or Gaussian

079. Which of the following are the two phases of RBF network learning? **A**

A Training and generalization
B Training and testing
C Training and validating
D Testing and validating

080. For strict interpolation, the interpolating surface is constrained to pass through ______. **A**

A All the training data points
B selected training data points
C All the testing data points
D selected testing data points

081. Which of the following is the common property shared by Gaussian and Inverse multiquadrics functions? **A**

A They are both localized functions
B They are both nonlocal functions
C They are both bell curve shaped
D There dont share any common property

082. Hyper surface reconstruction problem is said to be ill posed if _ **D**

A Existence, uniqueness and continuity conditions are all satisfied
B If any two of Existence, uniqueness and continuity conditions are satisfied
C If any one of Existence, uniqueness and continuity conditions is satisfied
D If any one of Existence, uniqueness and continuity conditions is not satisfied

083. ___ can be used for solving ill posed problems **A**

A Regularization
B Transformation
C Translation
D Standardization

084. Tikhonov &#39s regularization theory involves which of the following terms? **A**

A Standard error term
B Weighted error term
C Regularized error term
D Summative error term

085. Tikhonov &#39s standard error term measures the standard error between **A**

A Desired and actual response
B Input and output
C Induced local field and target output
D Desired response and target response

086. An ill posed problem can be made a well posed problem via __ **A**

A Regularization
B Transformation
C Translation
D Standardization

087. Automated vehicle is an example of ______. **A**

A Supervised learning
B Unsupervised learning
C Active learning
D Reinforcement learning

088. For which of the following functions, the interpolation matrix has N-1 negative eigen values and only one positive eigen value making it not positive definite? **A**

A Multiquadrics
B Inverse multiquadrics
C Inverse Quadratic
D Either Inverse Quadratic or Gausssian

089. Which of the following conditions is not needed for a reconstructing problem to be well posed. **C**

A Existence
B Uniqueness
C Linearity
D Continuity

090. Learning from physical forms of data such as speech, pictures, radar signals, viewed as a hyper surface reconstruction problem is ___ problem **D**

A Well posed direct
B Well posed inverse
C Ill posed direct
D Ill posed inverse

091. In practical applications, the regularization parameter is assigned a value somewhere between **B**

A -infinity and 0
B 0 and infinity
C -1 and 1
D 0 and 1

092. Regularization provides a practical solution to ______. **B**

A Biaslearningrate dilemma
B Bias-variance dilemma
C Learningrate-momentum dilemma
D Learningrate-variance dilemma

093. Frechet differential of a functional may be interpreted as the best _____ approximation. **A**

A Local linear
B Global linear
C Local non-linear
D Global non-linear

094. Tikhonov &#39s regularization theory involves which of the following terms? **A**

A Regularising term
B Weighted error term
C Regularized error term
D Summative error term

095. According to regularization theory, 0 implies **A**

A The solution can be completely determined by the examples
B The solution can be partially determined by the examples
C The solution cannot be completely determined by the examples
D The examples are unreliable

096. According to regularization theory, implies **C**

A The solution can be completely determined by the examples
B The solution can be partially determined by the examples
C Prior smoothness constraint is sufficient to specify the solution
D The examples are reliable

097. ____ proposed regularization method for solving ill posed problems. **C**

A Cover
B Powell
C Tikhonov
D Broomhead

098. The quantity to be minimized in regularization theory is called _____. **A**

A Tikhonov functional
B Miccheli functional
C Regularization parameter
D Minimization parameter

099. Regularization parameter may be viewed as an indicator of the **A**

A sufficiency of the given data set as examples that specify the solution F(X)
B sufficiency of the accuracy obtained for given set of examples
C insufficiency of the given data set as examples that specify the solution F(X)
D insufficiency of the accuracy obtained for given set of examples

100. Regularization prevents **B**

A Model under fitting
B Model over fitting
C Both over and under fitting
D Prediction error

101. Number of hidden nodes in regularization RBF network with N examples available for training is ___ **D**

A 1
B 2
C 3
D N

102. Which of the following parameters are unknown in the regularization RBF network? **D**

A Activation functions of the hidden layer
B Positions of the centres of the radial basis functions
C Norm weighting matrix associated with the hidden layer
D Linear weights of the output layer

103. Bellmans curse of dimensionality tells that irrespective of the approximation techniques **A**

employed, if the smoothness index is maintained constant the number of parameters needed for the approximating function to attain degree of accuracy increases __

A Exponentially with the input dimensionality
B Linearly with the input dimensionality
C constantly with the input dimensionality
D Polynomially with the input dimensionality

104. Which of the following is not a desirable property of regularization network? **D**

A It is a Universal approximator
B It has the best approximation property
C It computes optimal solution
D It is expensive

105. The condition number of a matrix is defined as __ **A**

A Ratio of largest eigen value to the smallest eigen value of the matrix
B Ratio of smallest eigen value to the largest eigen value of the matrix
C Ratio of largest value to the smallest value of the matrix
D Ratio of smallest value to the largest value of the matrix

106. Which of the following terms depend on the geometric properties of approximating function? **A**

A Regularising term
B Weighted error term
C Regularized error term
D Summative error term

107. ___ equation defines a necessary condition for the Tikhonov functional to have an extremum at F(x) **C**

A Greens function
B Dirac delta function
C Euler-Lagrange
D Hilbert

108. Greens function is a ____ function **A**

A Symmetric
B Unsymmetric
C Asymmetric
D Null

109. The space of approximating functions attainable with RBF networks becomes increasingly constrained as __ **A**

A Input dimensionality m_0 is increased
B Input dimensionality m_0 is decreased
C Number of hidden units are increased
D Number of output units are increased

110. The curse of dimensionality can be broken by __ **D**

A Either neural networks or any other non-linear technique of similar nature
B Multi layerperceptrons
C RBF networks
D Neither neural networks nor any other non-linear technique of similar nature

111. Activation function of each hidden unit in an RBF network computes __ **A**

A Euclidean norm
B Distance between input vector and output vector
C Outer product of the input vector and the synaptic weight vector of that unit
D Inner product of the input vector and the synaptic weight vector of that unit

112. Based on the linear characteristics of the output layer of RBF network, it is more closely associated to ___. **B**

A Multilayer perceptron
B Rosenblatts perceptron
C Either single layer perceptron or Multilayer perceptron
D Neither single layer perceptron nor Multilayer perceptron

113. To be immune to the curse of dimensionality, an approximating function should __ with the number of parameters. **A**

A Increase smoothness index
B Decrease smoothness index
C Increase number of training examples
D Decrease number of training examples

114. ____ method can be used to convert an ill posed problem into a well posed problem. **A**

A Regularization
B Transformation
C Translation
D Standardization

115. Model over fitting is prevented by __ **B**

A Normalization B Regularization
C Standardization D Neutralization

116. The ratio of largest eigen value to the smallest eigen value of the matrix is called __ number. **B**

A Eigen B Condition
C Singular D Cardinal

117. In regularization RBF network, the number of hidden nodes is __the number of examples available for training. **C**

A Less than B Greater than
C Equal to D No relation

118. Suppose you are using RBF kernel in SVM with high Gamma value. What does this signify? **B**

A The model would consider even far away points from hyperplane for modelling
B The model would consider only the points close to the hyperplane for modelling
C The model would not be affected by distance of points from hyperplane for modelling
D The model will be too constrained and include all points of the training dataset

119. The cost parameter in the SVM means: **C**

A The number of cross-validations to be made
B The kernel to be used
C The tradeoff between misclassification and simplicity of the model
D The size of input data

120. The effectiveness of an SVM depends upon: **D**

A Selection of Kernel
B Kernel Parameters
C Soft Margin Parameter C
D kernel, kernel parameters and soft margin parameter

121. If I am using all features of my dataset and I achieve 100% accuracy on my training set, but &#3970%on validation set, what should I look out for? **C**

A Underfitting B Nothing, the model is perfect
C Overfitting D Increase the training samples

122. What do you mean by a hard margin? **A**

A The SVM allows very low error in classification
B The SVM allows high amount of error in classification
C The SVM is very flexible in classification
D The SVM is generalized and works extremely well for unseen data

123. 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 Datasets which have a clear classification boundary

124. RBF networks construct ___ **D**

A Global approximations to linear input-output mapping
B Local approximations to linear input-output mapping
C Global approximations to non-linear input-output mapping
D Local approximations to non-linear input-output mapping

125. 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 the class labels of test data

126. Support vectors are the data points that lie _____. **A**

A Closest to the decision surface
B On the decision surface
C Farthest to the decision surface
D On either side of the decision surface

127. SVMs can be used for **A**

A Pattern Classification
B Pattern Association
C Pattern Clustering
D Pattern Storage

128. The main idea of SVM is to construct a hyperplane in such a way that the margin of separation between positive and negative examples is ______. **C**

A Minimized
B Neutralized
C Maximized
D Nullified

129. SVM is an approximate implementation of the method of **B**

A structural risk maximization
B structural risk minimization
C standard risk nullification
D standard risk minimization

130. The principle of SVM is based on the fact that generalization error rate is bounded by ________. **C**

A inner product of training error rate and a term that depends on VC dimension
B difference between training error rate and a term that depends on VC dimension
C sum of training error rate and a term that depends on VC dimension
D outer product of training error rate and a term that depends on VC dimension

131. What is the consequence of increasing the complexity (or degree of polynomial of the kernel used) linear kernel which perfectly fits the data? **A**

A Increasing the complexity will overfit the data
B Increasing the complexity will underfit the data
C Nothing will happen since the model was already 100% accurate
D Increases the misclassification rate for training data

132. SVM stands for **B**

A Standard Vector Machine
B Support Vector Machine
C Sequential Vector Model
D Structured Velocity Model

133. Which of the following is not an application of the SVM? **D**

A Text and Hypertext Categorization
B Image Classification
C Clustering of News Articles
D Payroll Generation

134. While training an SVM, increasing the number of features in training samples causes **D**

A Increased bias and variance
B Decreased bias and variance
C Increased bias and decreased variance
D Increased variance and decreased bias

135. Suppose you have trained an SVM with linear decision boundary after training SVM, you found that the model is under fitting. Which of the following options should you consider for resolving the problem? **C**

A increase the data points
B decrease the data points
C increase the features
D reduce the features

136. Optimal hyperplane is the hyperplane for which the margin of separation is______. **C**

A Minimized
B Neutralized
C Maximized
D Nullified

137. What is/are true about kernel in SVM? 1) Kernel function maps low dimensional data to high dimensional space 2) It is a similarity function **C**

A 1
B 2
C 1 and 2
D Neither 1 nor 2

138. In the context of using Gaussian kernel in SVM, which of the following statements is/are true about feature normalization? 1) We do feature normalization so that new feature will dominate other 2) Sometimes, feature normalization is not feasible in case of categorical variables 3) Feature normalization always helps when we use Gaussian kernel in SVM **B**

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

139. For an optimal hyperplane, the optimum condition is attained by minimizing the **A**

________.

A Euclidean norm of weight vector.
B Weight vector
C Number of training examples
D Number of features in a training vector

140. Given input vector x, weight vector w and bias b, the equation of a decision surface in the form of hyperplane that does the separation is_________. **B**

A $w^T x+b< 0$
B $w^T x+b = 0$
C $w^T x+b> 0$
D $w^T x+b>= 0$

141. Margin of separation is the separation between ______ **A**

A the hyperplane and the closest data point
B The hyperplane and the farthest data point
C any two closest data points
D any two farthest data points

142. In the case of separable patterns, SVM produces______ **B**

A a value of zero for both generalization error and the term that depends on VC dimension
B a value of zero for generalization error and minimizes the term that depends on VC dimension
C a value of zero for generalization error and maximizes the term that depends on VC dimension
D minimum values for both generalization error and the term that depends on VC dimension

143. The support vectors consist of __________. **A**

A subset of training data extracted by the algorithm
B subset of testing data extracted by the algorithm
C sum of the training examples calculated by the algorithm
D sum of the testing examples calculated by the algorithm

144. Support vector learning algorithm cant construct which of the following classifiers? **D**

A Polynomial learning machines
B Radial Basis Function networks
C Two-layer Perceptrons
D Bayesian classifiers

145. In the context of duality theorem, which of the following statements is true. **B**

A If the primal problem has an optimal solution, the dual problem also has an optimal solution, and the corresponding optimal values are not equal.
B If the primal problem has an optimal solution, the dual problem also has an optimal solution, and the corresponding optimal values are equal.
C If the primal problem has an optimal solution, the dual problem may not have an optimal solution, and the corresponding optimal values are equal.
D If the primal problem has an optimal solution, the dual problem may not have an optimal solution, and the corresponding optimal values are not equal.

146. VC dimension of a set of separating hyperplanes in a space of dimensionality m is equal to ______. **C**

A m
B m-1
C m+1
D -m

147. In a support vector machine a structure is imposed on the set of separating planes by constraining _______of the weight vector w **C**

A size
B sign
C Euclidean norm
D distribution

148. Optimal hyperplane is the separating hyperplane with _________. **A**

A largest margin of separation
B Smallest margin of separation
C Maximized cost function
D Maximized bias

149. The solution to the constrained optimization problem is determined by the saddle point of Lagrangian function, which has to be minimized with respect to **A**

A w and b
B w
C b
D w or b

150. The dual problem has the same optimal value as the primal problem, but with _______ providing the optimal solution. **D**

A Weight Vector B Bias
C Weight Vector and Bias D Lagrange multiplier

151. The constrained optimization problem is called _________. **B**

A Primary Problem B Primal Problem
C Dual Problem D Prime Problem

152. The constrained optimization problem may be solved by using the method of _____. **C**

A Gaussian Multipliers B Euclidean Multipliers
C Lagrange Multipliers D Bayesian Multipliers

153. The solution to the constrained optimization problem is determined by the saddle point of _______. **C**

A Gaussian function B Euclidean function
C Lagrangian function D Bayesian function

154. ________ measure deviation of a data point from the ideal condition of pattern separability. **A**

A Slack variables B Soft margin
C Hard margin D Euclidean norm

155. Minimization of () with respect to w is a __________optimization problem. **B**

A convex B nonconvex
C linear D polynomial

156. Nonconvex optimization problem is _________. **B**

A NP-Hard B NP-Complete
C NP D P

157. Support vector machine is a learning method for the design of a ____________. **A**

A feedforward network B feedback network
C recurrent network D Deep neural network

158. which of the following cases does not cause soft margin? **A**

A The data point (xi, di) falls outside the region of separation but on the right side of decision surface
B The data point (xi, di) falls inside the region of separation but on the right side of decision surface
C The data point (xi, di) falls outside the region of separation but on the wrong side of decision surface
D The data point (xi, di) falls inside the region of separation but on the wrong side of decision surface

159. Slack variables are ___________. **B**

A Non-negative vectors B Non-negative scalars
C Non-positive vectors D Non-positive scalars

160. Which of the following two statements is/are reason(s) for slower behaviour of SVM ? 1) There is no control over the number of data points selected by the learning algorithm for use as support vectors 2) There is no provision for incorporating prior knowledge about the task at hand into the design of learning machine **C**

A 1 B 2
C Both 1 and 2 D Neither 1 nor 2

161. If a set of non-separable patterns are given as training data, then the objective of learning algorithm is to find an optimal hyperplane that _________. **B**

A Doesnt encounter classification errors
B Minimizes the probability of classification errors
C Maximizes the probability of classification errors
D Either minimizes or maximizes the probability of classification errors

162. The margin of separation is said to be soft if a data point (xi,di) violates the condition ______ for i = .1, 2, .., N **D**

A $di(w^Txi+ b) > 0$ B $di(w^Txi+ b) >= 0$
C $di(w^Txi+ b) >+1$ D $di(w^Txi+ b) >= +1$

163. Which of the following two statements is true? 1) Backpropagation algorithm minimizes a quadratic loss function regardless of what the learning task is. 2) SVM also minimizes a quadratic loss function for reducing the misclassification rate. **A**

A 1 B 2
C Both 1 and 2 D Neither 1 nor 2

164. To reduce misclassification rate, support vector learning algorithm minimizes number of training samples that fall inside the margin of separation. This statement is approximately true because __. **B**

A Indicator function is used instead of slack variables B slack variables are used instead of indicator function
C slack variables are not used D Both slack variables and indicator function are used

165. Which of the following two statements is true? 1) Support vector learning algorithm minimizes number of training samples that fall inside the margin of separation to improve accuracy of classification. 2) Backpropagation algorithm minimizes a quadratic loss function regardless of what the learning task is **C**

A 1 B 2
C Both 1 and 2 D Neither 1 nor 2

166. SVM bypasses curse of dimensionality problem by focusing on the ______ for performing the constrained optimization problem. **B**

A primal problem B dual problem
C minimization of generalization error D maximization of classification accuracy

167. Which of the following are not included by SVMs? **D**

A Polynomial learning machine B RBF network
C Two-layer Perceptron D Single layer perceptron

168. SVM operates in **A**

A Batch mode B Sequential mode
C Pattern mode D Sample mode

169. SVM can be used to design a feedforward network with _____ hidden layer(s) of _______ units. **A**

A single, nonlinear B multiple, nonlinear
C single, linear D multiple, linear

170. Support vectors are **C**

A not part of training data B subset of test data
C subset of training data D average values computed from training data

171. SVM consists of quadratic programming problem, which of the following two reason(s) causes it attractive? 1) It is guaranteed to find a global extremum of the error surface 2) The computation can be performed efficiently. **C**

A 1 B 2
C Both 1 and 2 D Neither 1 nor 2

172. By using _______, SVM automatically finds all the important network parameters pertaining to that choice of a kernel **A**

A Inner-product kernel B Kernel type
C Training data D Test data

173. The memory requirement of the quadratic programming problem grows with _______. **B**

A Size of the training sample B Square of the size of the training sample
C Complexity of kernel D Number of target classes