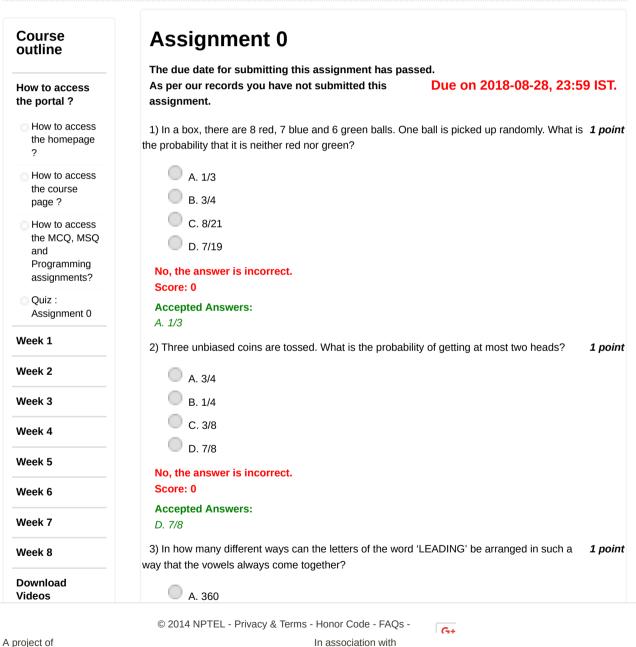
reviewer3@nptel.iitm.ac.in ▼

Courses » Introduction to Machine Learning

Announcements Course Ask a Question **Progress** Mentor FAQ

Unit 1 - How to access the portal?



A project of



Funded by

Government of India Ministry of Human Resource De

Accepted Answers: C. 720	
4) A bag contains 4 white, 5 red and 6 blue balls. Three balls are drawn at random from the bag. The probability that all of them are red, is:	1 point
A. 1/22	
B. 3/22	
C. 2/91	
O D. 2/77	
No, the answer is incorrect. Score: 0	
Accepted Answers: C. 2/91	
5) In a group of 6 boys and 4 girls, four children are to be selected. In how many different ways can they be selected such that at least one boy should be there?	1 point
A. 159	
О в. 194	
C. 205	
D. 209	
No, the answer is incorrect. Score: 0	
Accepted Answers: D. 209	
6) For a skew symmetric odd ordered matrix A of integers, which of the following will hold true:	1 point
A. det(A) = 9	
B. det(A) = 25	
C. det(A) = 0	
D. det(A) = 7	
No, the answer is incorrect. Score: 0	
Accepted Answers: C. det(A) = 0	
7) The function $f(x)=x^3-6x^2+9x+25$ has	1 point
\bigcirc A. a maxima at x= 1 and a minima at x = 3	
B. a maxima at $x = 3$ and a minima at $x = 1$	
\bigcirc C. no maxima, but a minima at x = 1	
\bigcirc D. a maxima at x = 1, but no minima	
No, the answer is incorrect.	

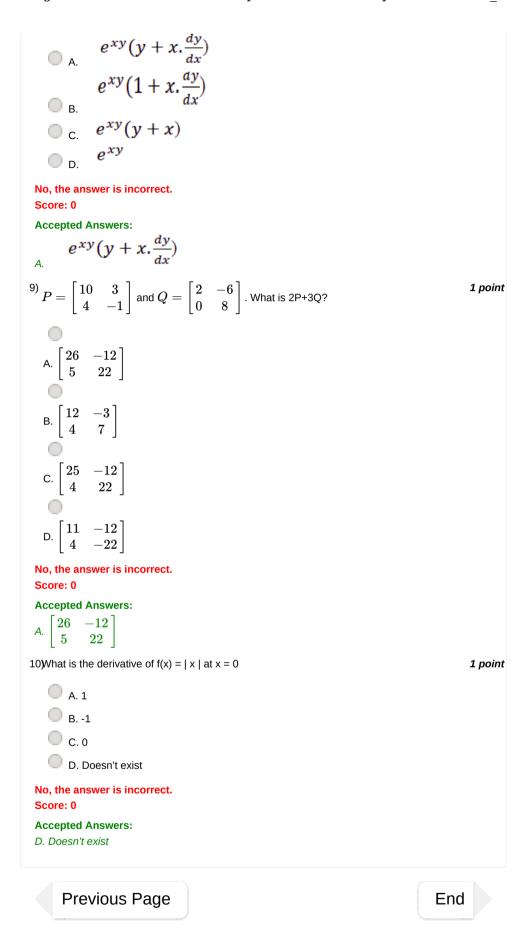
Score: 0

Accepted Answers:

A. a maxima at x = 1 and a minima at x = 3

$$\frac{d}{dx}(e^{xy}) = ?$$

1 point



Χ

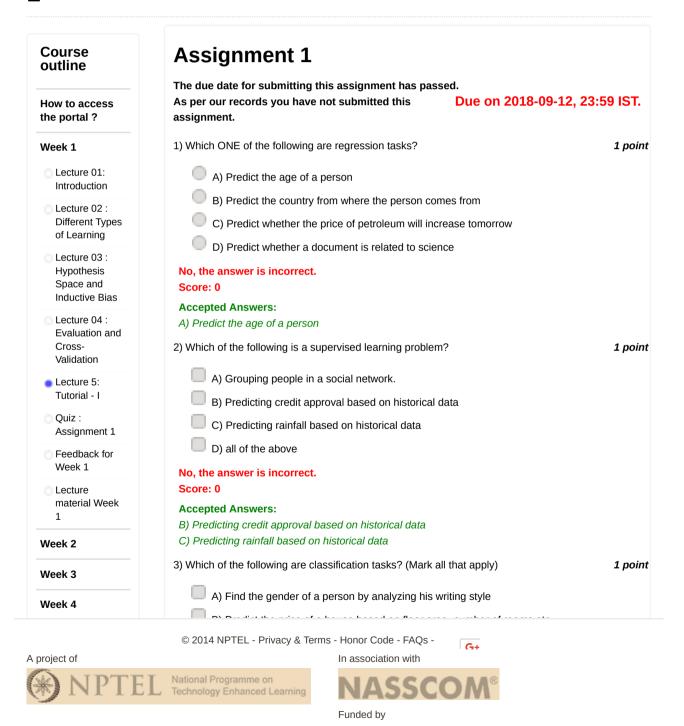
NIPTEIL

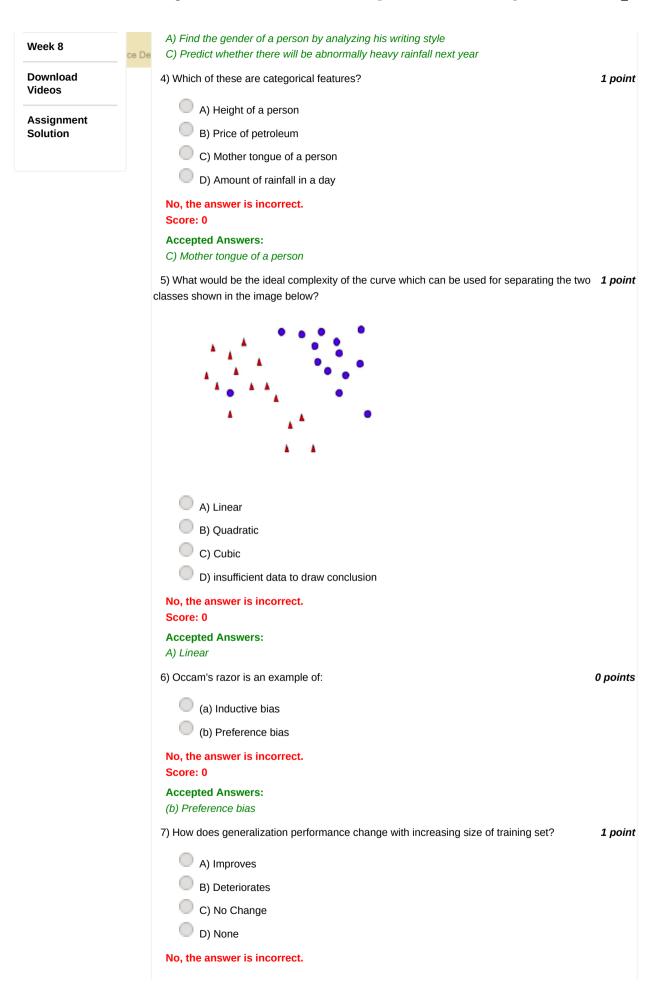
reviewer3@nptel.iitm.ac.in ▼

Courses » Introduction to Machine Learning

Announcements Course Ask a Question Progress Mentor FAQ

Unit 2 - Week 1





Score: 0

Accepted Answers:

A) Improves

8) One of the most common uses of Machine Learning today is in the domain of Robotics. **1 point** Robotic tasks include a multitude of ML methods tailored towards navigation, robotic control and a number of other tasks. Robotic control includes controlling the actuators available to the robotic system. An example of this is control of a painting arm in automotive industries. The robotic arm must be able to paint every corner in the automotive parts while minimizing the quantity of paint wasted in the process. Which of the following learning paradigms would you select for training such a robotic arm?

A) Supervised learning

B) Unsupervised learning

C) Combination of supervised and unsupervised learning

D) Reinforcement learning

No, the answer is incorrect.

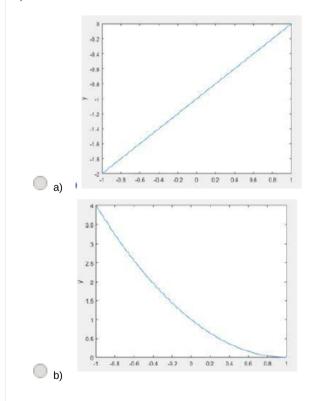
Score: 0

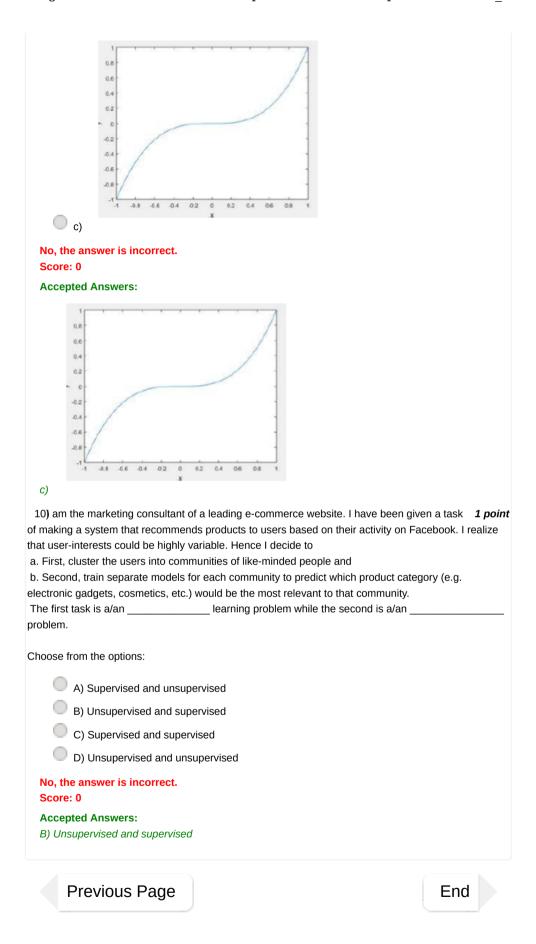
Accepted Answers:

D) Reinforcement learning

9) Choose the function that has the maximum variance:

1 point





Progress

FAQ

Mentor



Course

Ask a Question

Announcements

Unit 3 - Week 2

Course outline	Assignment 2	
How to access the portal ?	The due date for submitting this assignment has passed. As per our records you have not submitted this assignment.	Due on 2018-09-12, 23:59 IST.
portal ?	1) In regression the output is	1 point
Week 1	A) Discrete.	
Week 2	B) Continuous and always lies in a finite range.	
Lecture 06 :	C) Continuous.	
Linear Regression	D) May be discrete or continuous.	
Lecture 07 : Introduction to Decision Trees	No, the answer is incorrect. Score: 0	
Lecture 08 :	Accepted Answers:	
Learning Decision Tree	C) Continuous.	d water
Lecture 09 : Overfitting	2) In linear regression the parameters are A) strictly integers	1 point
Lecture 10: Python Exercise on Decision Tree and Linear Regression	B) always lies in the range [0,1]C) any value in the real spaceD) any value in the complex space	
Lecture 11: Tutorial - II	No, the answer is incorrect. Score: 0	
Lecture notes - Week 2	Accepted Answers: C) any value in the real space	
Quiz : Assignment 2	3) Which of the following is true for a decision tree?	0 points
Feedback for	A) Decision tree is an example of linear classifier.	
Week 2	B) The entropy of a node typically decreases as we go do	wn a decision tree.
Week 3	C) Entropy is a measure of purity.	
Neek 4	D) An attribute with lower mutual information should be pro	eferred to other attributes.
Week 5	No, the answer is incorrect. Score: 0	
Week 6	Accepted Answers:	

© 2014 NPTEL - Privacy & Terms - Honor Code - FAQs
A project of

In association with

NATIONAL Programme on
Technology Enhanced Learning

Funded by

Monday 05 November 2018 04:22 PM

Powered by

Solution

C) 0.50

D) 0.22

No, the answer is incorrect.
Score: 0

Accepted Answers:

A) 0.940

5)

1 point

Outlook	Temperature	Humidity	Wind	Play tennis
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Overcast	Hot	High	Weak	Yes
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
overcast	Cool	Normal	Strong	Yes
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Rain	Mild	Normal	Weak	Yes
Sunny	Mild	Normal	Strong	Yes
Overcast	Mild	High	Strong	Yes
Overcast	Hot	Normal	Weak	Yes
Rain	Mild	High	strong	No

The decision on whether tennis can be played or not is based on the following features: Outlook E {Sunny, Overcast, Rain}, Temperature E {Hot, Mild, Cool}, Humidity E {High, Normal} and Wind E {Weak, Strong}. The training data is given above.

The entropy of the entire dataset is

A) 1

B) 0.94

C) 0

O D) 0.72

No, the answer is incorrect.

Score: 0

Accepted Answers:

B) 0.94

6) 1 point

Outlook	Temperature	Humidity	Wind	Play tennis
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Overcast	Hot	High	Weak	Yes
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
overcast	Cool	Normal	Strong	Yes
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Rain	Mild	Normal	Weak	Yes
Sunny	Mild	Normal	Strong	Yes
Overcast	Mild	High	Strong	Yes
Overcast	Hot	Normal	Weak	Yes
Rain	Mild	High	strong	No

The decision on whether tennis can be played or not is based on the following features: Outlook E {Sunny,

Overcast, Rain}, Temperature E {Hot, Mild, Cool}, Humidity E {High, Normal} and Wind E {Weak, Strong}. The training data is given above.
Which attribute will be the root of the decision tree and how much is the information gain due to the attribute.
A) Outlook, 0.246
B) Humidity, 0.5
C) Temperature, 0.306
D) Humidity, 0.48
No, the answer is incorrect.
Score: 0
Accepted Answers:
A) Outlook, 0.246

7) ISRO wants to discriminate between Martians (M) and Humans (H) based on the following features: **1** point Green E {N,Y}, Legs E {2,3}, Height E {S,T},Smelly E {N,Y}. The training data is as follows:

Species	Green	Legs	Height	Smelly
M	N	3	S	Υ
М	Y	2	Т	N
М	Υ	3	Т	N
М	N	2	S	Y
М	Y	3	Т	N
Н	N	2	Т	Y
Н	N	2	S	N
Н	N	2	Т	N
Н	Y	2	S	N
Н	N	2	T	Y

int

Species	Green	Legs	Height	Smelly
M	N	3	S	Y
М	Y	2	Т	N
М	Υ	3	Т	N
М	N	2	S	Y
М	Υ	3	Т	N
Н	N	2	Т	Y
Н	N	2	S	N
Н	N	2	Т	N
Н	Y	2	S	N
Н	N	2	T	Y

how much is the information gain due to the attribute found in the previous question?

A) 0.45
B) 0.40
C) 0.80
D) 0.70

No, the answer is incorrect.
Score: 0

Accepted Answers:

9) The following table shows the results of a recently conducted study on the correlation of the number **1** point of hours spent driving with the risk of developing acute back-ache. Find the equation of the best fit line for this data.

Number of hours spent driving (x)	Risk score on a scale of 0-100 (y)
10	95
9	80
2	10
15	50
10	45
16	98
11	38
16	93

- A) y = 3.39x + 11.62
- B) Y = 4.69x + 12.58
- C) Y = 4.59x + 12.58
- D) Y = 3.59x + 10.58

No, the answer is incorrect.

Score: 0

B) 0.40

Accepted Answers:

C) Y = 4.59x + 12.58

10)Decision trees can be used for the following type of datasets:

I. The attributes are categorical

1 point

II. The attributes are numeric valued and continuous		
III. The attributes are discrete valued numbers		
A) In case I only B) In case II only C) In casesII and III only D) In cases I, II and III		
No, the answer is incorrect. Score: 0		
Accepted Answers:		
D) In cases I, II and III		
Previous Page		End
	A) In case I only B) In case II only C) In casesII and III only D) In cases I, II and III No, the answer is incorrect. Score: 0 Accepted Answers: D) In cases I, II and III	III. The attributes are discrete valued numbers A) In case I only B) In case II only C) In casesII and III only D) In cases I, II and III No, the answer is incorrect. Score: 0 Accepted Answers: D) In cases I, II and III



reviewer3@nptel.iitm.ac.in ▼

Courses » Introduction to Machine Learning

Announcements Course Ask a Question Progress Mentor FAQ

Unit 4 - Week 3

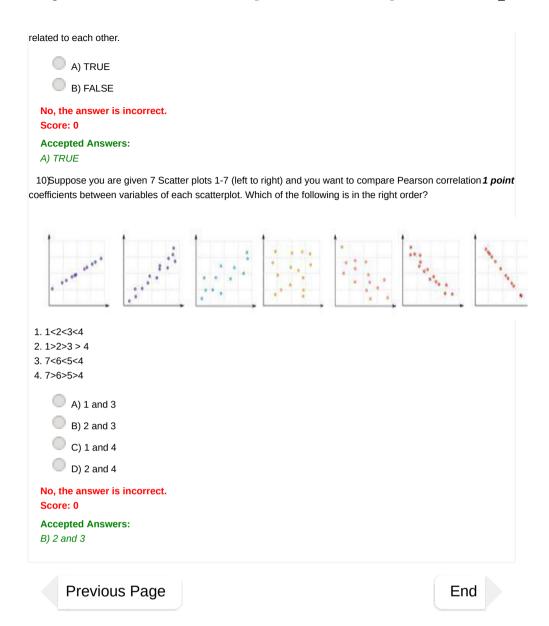
Course outline	Assignment 3	
How to access the	The due date for submitting this assignment has passed. As per our records you have not submitted this assignment.	59 IST.
portal ?	1) In k-NN algorithm, given a set of training examples and the value of k < size of training set (n), the	1 point
Week 1	algorithm predicts the class of a test example to be the	1 point
Week 2	A) Most frequent class among the classes of k closest training examples.	
Week 3	B) Least frequent class among the classes of k closest training examples.	
Lecture 12:	C) Class of the closest point.	
k-Nearest Neighbour	D) Most frequent class among the classes of the k farthest training examples.	
	No, the answer is incorrect.	
Lecture 13: Feature Selection	Score: 0	
Lecture 14:	Accepted Answers:	
Feature	A) Most frequent class among the classes of k closest training examples.	
Extraction	2) In collaborative Filtering based Recommendation, the items are recommended based on which of	0 points
Lecture 15: Collaborative	the following?	
Filtering	A) Similar users	
Lecture 16:	B) Similar items	
Python Exercise on kNN and PCA	C) Both A and B	
Lecture 17:	D) None	
Tulonarm	No, the answer is incorrect.	
Lecture notes - Week 3	Score: 0	
	Accepted Answers:	
Quiz : Assignment	A) Similar users	
Feedback for	3) Which of the following are advantages of large value of k in k-NN algorithm?	1 point
Week 3	A) Less sensitive to noise.	
Week 4	B) Better probability estimates for discrete classes.	
Week 5	C) Larger training sets allow larger values of .	
Week 6	D) All of the above.	
	No, the answer is incorrect.	
Week 7	Score: 0	

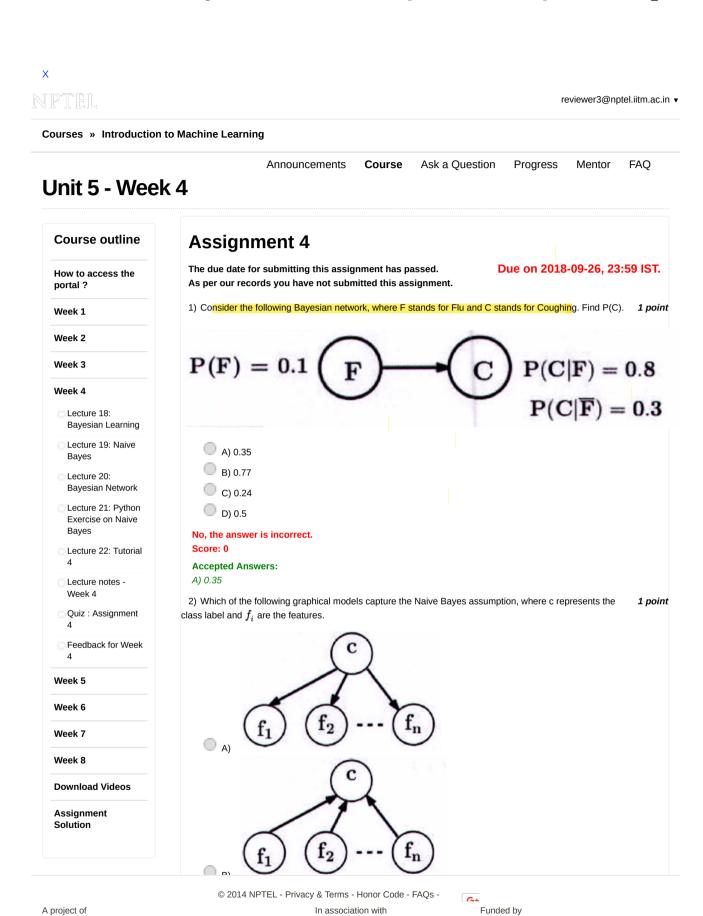
© 2014 NPTEL - Privacy & Terms - Honor Code - FAQs A project of In association with

NATIONAL PROGRAMME ON Technology Enhanced Learning
Funded by

Powered by

ministry or numan resource Devel	C) Limited computational resources.	
	D) All of the above.	
	No, the answer is incorrect. Score: 0	
	Accepted Answers: D) All of the above.	
	5) For which of the following cases Dimensional reduction may be used?	1 point
	A) Data Compression	
	B) Data Visualization	
	C) To prevent overfitting	
	D) Both A and B	
	No, the answer is incorrect.	
	Score: 0	
	Accepted Answers: D) Both A and B	
	6) Which of the following is the limitation of Collaborative Filtering?	1 point
	A) Over specialization	
	B) Cold start	
	C) Both A and B	
	D) None	
	No, the answer is incorrect. Score: 0	
	Accepted Answers: B) Cold start	
	7) Which of the following statements is true about PCA? (i) We must standardize the data before applying PCA. (ii) We should select the principal components which explain the highest variance (iii) We should select the principal components which explain the lowest variance (iv) We can use PCA for visualizing the data in lower dimensions	1 point
	A. (i), (ii) and (iv) B. (ii) and (iv) C. (iii) and (iv) D. (i) and (iii) No, the answer is incorrect. Score: 0 Accepted Answers: A. (i), (ii) and (iv)	
	8) In feature selection, which of the following techniques can be used to find a subset of features?	1 point
	A) Sequential forward search	
	B) Sequential backward search	
	C) Both A and B	
	D) None of A or B	
	No, the answer is incorrect. Score: 0	
	Accepted Answers: C) Both A and B	
	9) [True or False] A Pearson correlation between two variables is zero but, still their values can still be	1 point





National Programme on Technology Enhanced Learning

Powered by

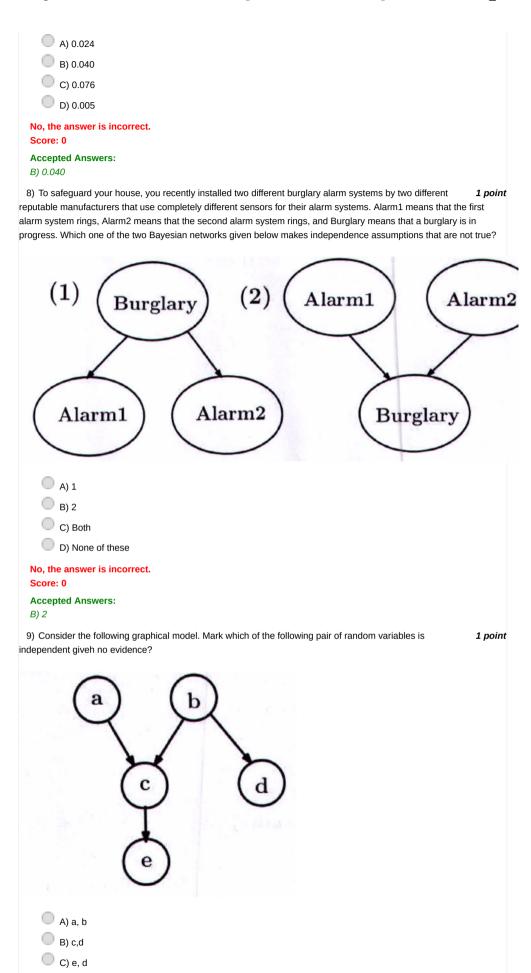
1 of 4

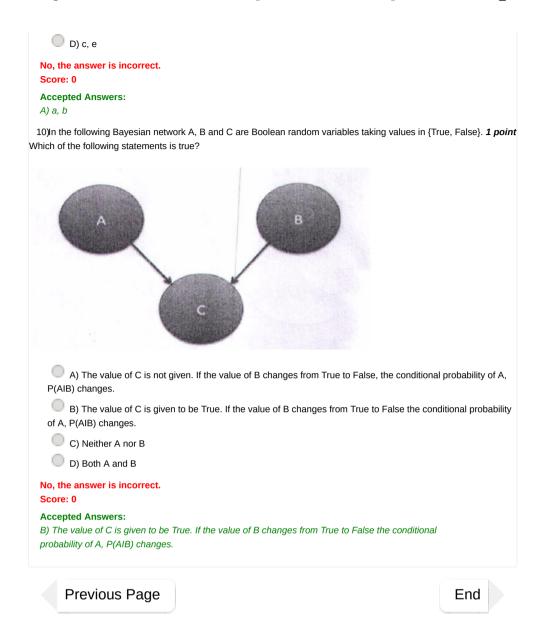
Government of India Ministry of Human Resource Developm



$\stackrel{\mathbf{c}}{\bigcirc}$
f_1 f_2 \cdots f_n
A)3) Bayesian Network is a graphical model that efficiently encodes the joint probability distribution for a large 1 point set of variables .
A) True B) False
No, the answer is incorrect. Score: 0
Accepted Answers: A) True
4) A fair coin is tossed three times and a T (for tails) or H (for heads) is recorded, giving us a list of length 3. 1 point Let X be the random variable which is zero if no T has another T adjacent to it, and is one othetwise. Let Y denote the random variable that counts the number of T's in the three tosses. Find P(X=1, Y=2).
A) 1/8
B) 2/8
C) 5/8
O D) 7/8
No, the answer is incorrect. Score: 0
Accepted Answers: B) 2/8
5) Two cards are drawn at random from a deck of 52 cards without replacement. What is the probability of drawing a 2 and an Ace in that order?
A) 4/51
B) 1/13
C) 4/256
O D) 4/663
No, the answer is incorrect. Score: 0
Accepted Answers: D) 4/663
6) A and B throw alternately a pair of dice. A wins if he throws 6 before B throws 7 and B wins if she throws 7 1 point before A throws 6. If A begins, his chance of winning would be:
A) 30/61
В) 31/61
C) 1/2
O D) 6/7
No, the answer is incorrect. Score: 0
Accepted Answers: A) 30/61
7) Diabetic Retinopathy is a disease that affects 80% people who have diabetes for more than 10 years. 5% 1 point of the Indian population has been suffering from diabetes for more than 10 years. Answer the following questions. What is the joint probability of finding an Indian suffering from Diabetes for more than 10 years and also has Diabetic

Retinopathy?





Х

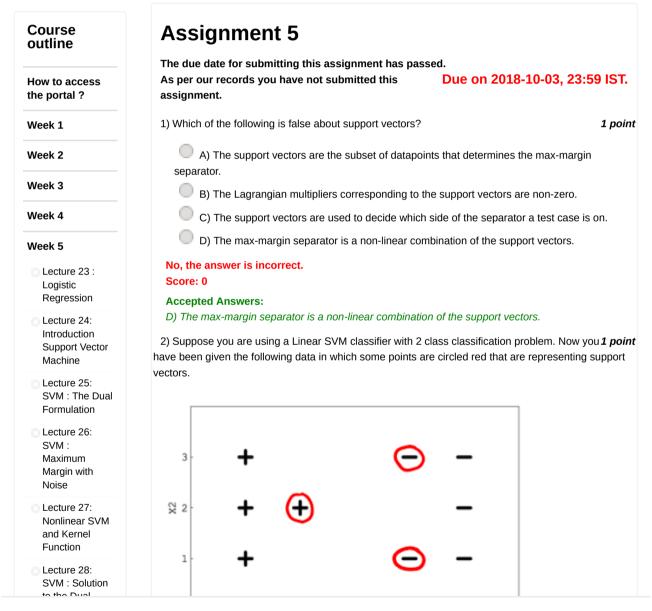
NIPTEI

reviewer3@nptel.iitm.ac.in ▼

Courses » Introduction to Machine Learning

Announcements Course Ask a Question Progress Mentor FAQ

Unit 6 - Week 5



© 2014 NPTEL - Privacy & Terms - Honor Code - FAQs -

G+

A project of





Funded by

In association with

Quiz: Assignment 5 Feedback for Week 5 Week 6 Week 7 Week 8 Download Videos Assignment Solution Assignment Solution Assignment Solution Assignment Solution No, the answer is incorrect. Score: 0 Accepted Answers: A) Yes 3) Consider a binary classification problem. Suppose I have trained a model on a linearl separable training set, and now I get a new labeled data point which is correctly classified model, and far away from the decision boundary. If I now add this new point to my earlier and re-train, in which cases is the learnt decision boundary likely to change? A) When my model is a perceptron. B) When my model is logistic regression. C) When my model is Gaussian discriminant analysis. No, the answer is incorrect. Score: 0 Accepted Answers: B) When my model is logistic regression. D) When my model is Gaussian discriminant analysis. 4) After training an SVM, we can discard all examples which do not support vectors and still classify new examples? A) TRUE B) FALSE No, the answer is incorrect. Score: 0	
Score: 0 Accepted Answers: A) Yes Week 6 Week 7 Week 8 Download Videos Assignment Solution Assignment Solution D) When my model is Gaussian discriminant analysis. No, the answer is incorrect. Score: 0 Accepted Answers: A) Yes 3) Consider a binary classification problem. Suppose I have trained a model on a linearl separable training set, and now I get a new labeled data point which is correctly classified model, and far away from the decision boundary. If I now add this new point to my earlier and re-train, in which cases is the learnt decision boundary likely to change? A) When my model is a perceptron. C) When my model is logistic regression. D) When my model is Gaussian discriminant analysis. No, the answer is incorrect. Score: 0 Accepted Answers: B) When my model is Gaussian discriminant analysis. 4) After training an SVM, we can discard all examples which do not support vectors and still classify new examples? A) TRUE B) FALSE No, the answer is incorrect.	
Week 5 Week 6 Week 7 Week 8 Download Videos Assignment Solution D) When my model is a perceptron. C) When my model is Gaussian discriminant analysis. No, the answer is incorrect. Score: 0 Accepted Answers: A) Yes 3) Consider a binary classification problem. Suppose I have trained a model on a linearly separable training set, and now I get a new labeled data point which is correctly classified model, and far away from the decision boundary. If I now add this new point to my earlier and re-train, in which cases is the learnt decision boundary likely to change? A) When my model is a perceptron. C) When my model is logistic regression. D) When my model is Gaussian discriminant analysis. No, the answer is incorrect. Score: 0 Accepted Answers: B) When my model is Gaussian discriminant analysis. 4) After training an SVM, we can discard all examples which do not support vectors and still classify new examples? A) TRUE B) FALSE No, the answer is incorrect.	
Week 6 Week 7 Week 8 Download Videos Assignment Solution On the answer is incorrect. By When my model is logistic regression. Dy When my model is logistic regression. Dy When my model is logistic regression. Accepted Answers: By When my model is logistic regression. Dy When my model is logistic regression. Accepted Answers: By When my model is logistic regression. Dy When my model is logistic regression. Accepted Answers: By When my model is logistic regression. Dy When my model is Gaussian discriminant analysis. 4) After training an SVM, we can discard all examples which do not support vectors and still classify new examples? A) TRUE B) FALSE No, the answer is incorrect.	
Week 7 Week 8 Download Videos Assignment Solution D) When my model is Gaussian discriminant analysis. No, the answer is incorrect. Score: 0 Accepted Answers: B) When my model is logistic regression. D) When my model is logistic regression. Accepted Answers: B) When my model is logistic regression. Accepted Answers: B) When my model is logistic regression. Accepted Answers: B) When my model is logistic regression. Accepted Answers: B) When my model is logistic regression. D) When my model is Gaussian discriminant analysis. 4) After training an SVM, we can discard all examples which do not support vectors and still classify new examples? A) TRUE B) FALSE No, the answer is incorrect.	
Assignment Solution D) When my model is Gaussian discriminant analysis. No, the answer is incorrect. Score: 0 Accepted Answers: B) When my model is logistic regression. D) When my model is Gaussian discriminant analysis. 4) After training an SVM, we can discard all examples which do not support vectors and still classify new examples? A) TRUE B) FALSE No, the answer is incorrect.	
Assignment Solution D) When my model is logistic regression. D) When my model is Gaussian discriminant analysis. No, the answer is incorrect. Score: 0 Accepted Answers: B) When my model is logistic regression. D) When my model is Gaussian discriminant analysis. 4) After training an SVM, we can discard all examples which do not support vectors and still classify new examples? A) TRUE B) FALSE No, the answer is incorrect.	training set
Assignment Solution D) When my model is logistic regression. D) When my model is Gaussian discriminant analysis. No, the answer is incorrect. Score: 0 Accepted Answers: B) When my model is logistic regression. D) When my model is Gaussian discriminant analysis. 4) After training an SVM, we can discard all examples which do not support vectors and still classify new examples? A) TRUE B) FALSE No, the answer is incorrect.	
Assignment Solution C) When my model is an SVM. D) When my model is Gaussian discriminant analysis. No, the answer is incorrect. Score: 0 Accepted Answers: B) When my model is logistic regression. D) When my model is Gaussian discriminant analysis. 4) After training an SVM, we can discard all examples which do not support vectors and still classify new examples? A) TRUE B) FALSE No, the answer is incorrect.	
D) When my model is Gaussian discriminant analysis. No, the answer is incorrect. Score: 0 Accepted Answers: B) When my model is logistic regression. D) When my model is Gaussian discriminant analysis. 4) After training an SVM, we can discard all examples which do not support vectors and still classify new examples? A) TRUE B) FALSE No, the answer is incorrect.	
No, the answer is incorrect. Score: 0 Accepted Answers: B) When my model is logistic regression. D) When my model is Gaussian discriminant analysis. 4) After training an SVM, we can discard all examples which do not support vectors and still classify new examples? A) TRUE B) FALSE No, the answer is incorrect.	
Score: 0 Accepted Answers: B) When my model is logistic regression. D) When my model is Gaussian discriminant analysis. 4) After training an SVM, we can discard all examples which do not support vectors and still classify new examples? A) TRUE B) FALSE No, the answer is incorrect.	
 B) When my model is logistic regression. D) When my model is Gaussian discriminant analysis. 4) After training an SVM, we can discard all examples which do not support vectors and still classify new examples? A) TRUE B) FALSE No, the answer is incorrect. 	
 D) When my model is Gaussian discriminant analysis. 4) After training an SVM, we can discard all examples which do not support vectors and still classify new examples? A) TRUE B) FALSE No, the answer is incorrect. 	
4) After training an SVM, we can discard all examples which do not support vectors and still classify new examples? A) TRUE B) FALSE No, the answer is incorrect.	
still classify new examples? A) TRUE B) FALSE No, the answer is incorrect.	
B) FALSE No, the answer is incorrect.	I can 1 point
No, the answer is incorrect.	
Accepted Answers: A) TRUE	
5) If $g(z)$ is the sigmoid function, then its derivative with respect to z may be written in te $g(z)$ as	erm of 1 point
A) g(z)(1-g(z))	
B) g(z)(1+g(z))	
C) $-g(z)(1+g(z))$	
D) g(z)(g(z)-1)	
No, the answer is incorrect. Score: 0	
Accepted Answers:	
A) g(z)(1-g(z))	
6) Which of the following are true when comparing ANNs and SVMs?	1 point
A) ANN error surface has multiple local minima while SVM error surface has on	ly one minima
B) After training, an ANN might land on a different minimum each time, when in random weights during each run.	itialized with
C) In training, ANN's error surface is navigated using a gradient descent technic SVM's error surface is navigated using convex optimization solvers.	que while
D) As shown for Perceptron, there are some classes of functions that cannot be ANN. An SVM can learn a hyperplane for any kind of distribution.	que vime

No, the answer is incorrect.

Score: 0

Accepted Answers:

- A) ANN error surface has multiple local minima while SVM error surface has only one minima
- B) After training, an ANN might land on a different minimum each time, when initialized with random weights during each run.
- C) In training, ANN's error surface is navigated using a gradient descent technique while SVM's error surface is navigated using convex optimization solvers.
- 7) Which of the following is not a kernel function?

1 point

https://onlinecourses.nptel.ac.in/noc18 cs40/uni...

- $A) K(x_i, x_j) = x_i. x_j$
- B) $K(x_i, x_j) = (1 x_i \cdot x_j)^3$
- C) $K(x_i, x_j) = e(-\|x_i x_j\|^2 / (2\sigma^2))$
- D) $K(x_i, x_j) = tanh(\beta 0 x_i. x_j + \beta 1)$

No, the answer is incorrect.

Score: 0

Accepted Answers:

B)
$$K(x_i, x_j) = (1 - x_i \cdot x_j)^3$$

- 8) Which of the following is true about SMO algorithm (multiple answers)?
- 1 point

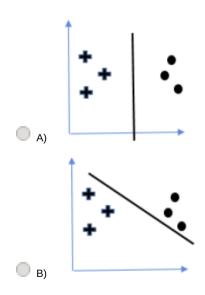
- A) The SMO can efficiently solve the primal problem.
- B) The SMO can efficiently solve the dual problem
- C) The SMO solves the optimization problem by co-ordinate ascent.
- D) The SMO solves the optimization problem by co-ordinate descent.

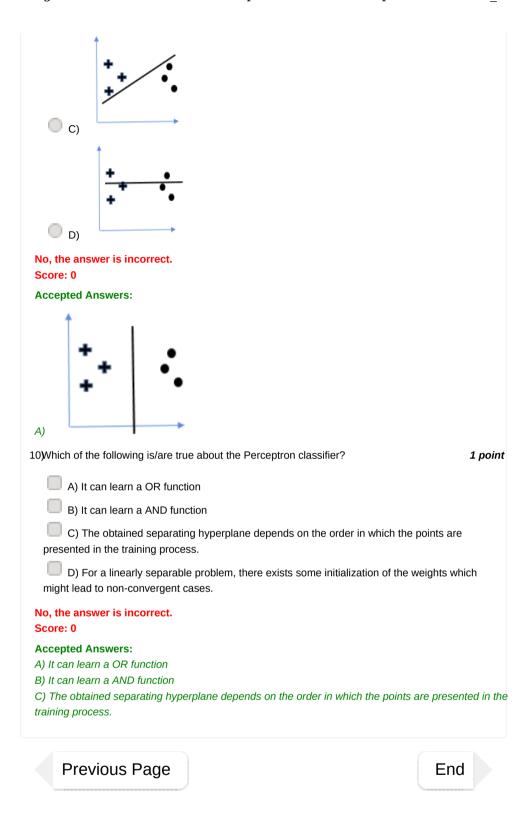
No, the answer is incorrect.

Score: 0

Accepted Answers:

- B) The SMO can efficiently solve the dual problem
- C) The SMO solves the optimization problem by co-ordinate ascent.
- 9) For the given set of points which of the following lines is most suitable to be the decision **1** point boundary?





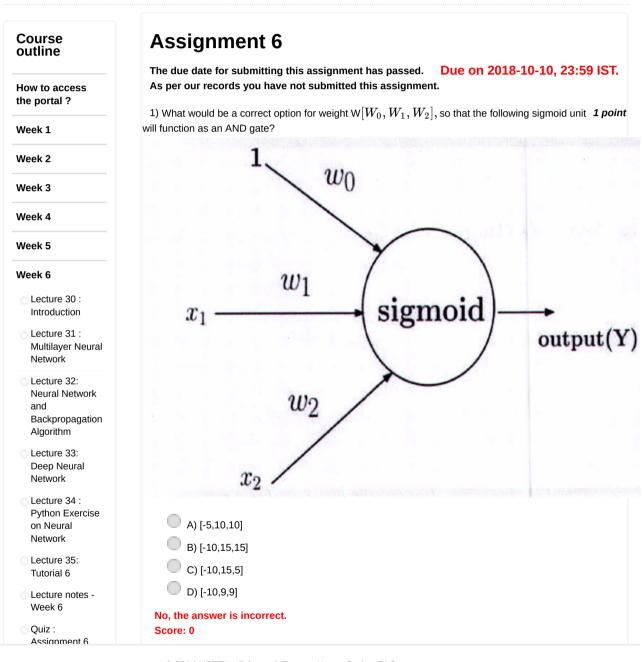


reviewer3@nptel.iitm.ac.in ▼

Courses » Introduction to Machine Learning

Announcements Course Ask a Question Progress Mentor FAQ

Unit 7 - Week 6



© 2014 NPTEL - Privacy & Terms - Honor Code - FAQs -

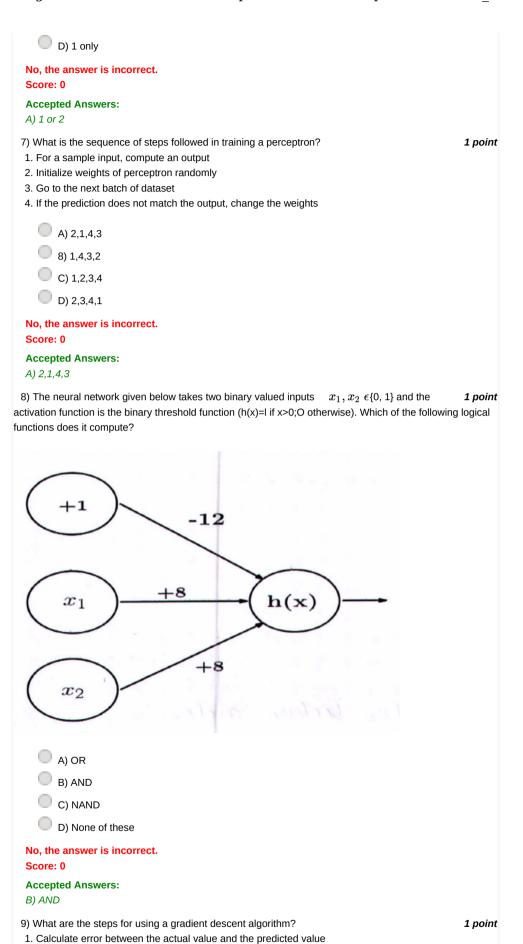




G+

Funded by

Download Videos	C) never finds the globally optimal solution.
Download videos	D) finds a locally optimal solution which is never globally optimal
Assignment Solution	No, the answer is incorrect.
	Score: 0
	Accepted Answers:
	B) finds a locally optimal solution which may be globally optimal.
	3) Which of the following is true? 1 point
	A) In batch gradient descent we update the weights and biases of the neural network after forward pass over each training example.
	B) In batch gradient descent we update the weights and biases of our neural network after forward pass over all the training examples.
	C) Each step of stochastic gradient descent takes more time than each step of batch gradient descent.
	D) None of these three options is correct
	No, the answer is incorrect. Score: 0
	Accepted Answers: B) In batch gradient descent we update the weights and biases of our neural network after forward pass over all the training examples.
	4) In a neural network, which one of the following techniques is NOT useful to reduce overfitting? 1 point
	A) Dropout
	B) Regularization
	C) Batch normalizatioh
	D) Adding more layers
	No, the answer is incorrect. Score: 0
	Accepted Answers: D) Adding more layers
	5) For an image recognition problem (such as recognizing a cat in a photo), which architecture of 1 point neural network has been found to be better suited for the tasks?
	A) Multi layer perceptron
	B) Recurrent neural network
	C) Convolutional neural network
	D) Perceptron
	No, the answer is incorrect. Score: 0
	Accepted Answers: C) Convolutional neural network
	 6) In training a batch neural network, after running the first few epochs, you notice that the loss does not decrease. The reasons for this could be 1. The learning rate is low. 2. The neural net is stuck in local minima 3. The neural net has too many units in the hidden layer
	A) 1 or 2
	B) 1 or 3
	C) 2 or 3



2. Repeat until you find the best weights of network

3. Pass an input through the network and get values from output layer

4. Initialize random values for weight and bias5. Go to each neurons which contributes to the error and change its respective values to reduce the
error
A) 4,3,1,5,2 B) 1,2,3,4,5 C) 3,4,5,2,1 D) 2,3,4,5,1 No, the answer is incorrect.
Score: 0 Accepted Answers: A) 4,3,1,5,2
10)A 4-input neuron has bias of 0 and weights 1, 2, 3 and 4. The transfer function is given by $f(v) = 1$ point $max(O,v)$. The inputs are 4, 10, 5 and 20 respectively. The output will be
O A) 238
О в) 119
C) 75
O D) 121
No, the answer is incorrect. Score: 0 Accepted Answers: B) 119
Previous Page End

Х

NIPTEL

reviewer3@nptel.iitm.ac.in ▼

Courses » Introduction to Machine Learning

Announcements Course Ask a Question Progress Mentor FAQ

Unit 8 - Week 7

Course outline	Assignment 7
How to access the portal ?	The due date for submitting this assignment has passed. As per our records you have not submitted this assignment. Due on 2018-10-17, 23:59 IST. assignment.
Week 1	1) The VC dimension of hypothesis space H1 is larger than the VC dimension of hypothesis 1 <i>point</i> space H2. Which of the following can be inferred from this?
Week 2	A) The number of examples required for learning a hypothesis in H1 is larger than the
Week 3	number of examples required for H2.
Week 4	B) The number of examples required for learning a hypothesis in H1 is smaller than the number of examples required for H2.
Week 5	C) No relation to number of samples required for PAC learning.
Week 6	No, the answer is incorrect. Score: 0
Week 7	Accepted Answers:
Lecture 36 :	A) The number of examples required for learning a hypothesis in H1 is larger than the number of examp required for H2.
Computational Learning Theory	2) In ensemble learning, you aggregate the predictions for weak learners, so that an 0 points ensemble of these models will give a better prediction than prediction of individual models. Which of
Lecture 37 :	the following statements is / are true for weak learners used in ensemble model? 1. They don't usually overfit.
Sample Complexity : Finite Hypothesis Space	2. They have high bias, so they cannot solve complex learning problems 3. They usually overfit. A) 1 and 2
Lecture 38: VC	B) 1 and 3
Lecture 39: Introduction to Ensembles	C) 2 and 3 D) Only 1
Lecture 40:	No, the answer is incorrect. Score: 0

© 2014 NPTEL - Privacy & Terms - Honor Code - FAQs -

G+

A project of





Funded by

In association with

Feedback for	C) is the hypothesis that gives best result on test instances.
Week 7	D) none of the above
Week 8	No, the answer is incorrect.
Download Videos	Score: 0
	Accepted Answers: B) is an ensemble of all the hypotheses in the hypothesis space.
Assignment Solution	4) For a particular learning task, if the requirement of error parameter ϵ changes from 0.1 to 1 point 0.01. How many more samples will be required for PAC learning?
	A) Same
	B) 2 times
	C) 10 times
	D) 1000 times
	No, the answer is incorrect.
	Score: 0
	Accepted Answers: C) 10 times
	5) Data scientists always use multiple algorithms for prediction and they combine output of 1 point multiple machine learning algorithms (known as "Ensemble Learning") for getting more robust or generalized output which outperform all the individual models. In which of the following options you think this is true (Choose best possible answer)?
	A) Base models having the higher correlation.
	B) Base models having the lower correlation.
	C) Use "Weighted average" instead of "Voting" methods of ensemble.
	D) Base models coming from the same algorithm
	No, the answer is incorrect. Score: 0
	Accepted Answers:
	B) Base models having the lower correlation.
	6) Suppose the VC dimension of a hypothesis space is 4. Which of the following are true? 1 point
	A) No sets of 4 points can be shattered by the hypothesis space.
	B) Atleast one set of 4 points can be shattered by the hypothesis space.
	C) All sets of 4 points can be shattered by the hypothesis space.
	D) No set of 5 points can be shattered by the hypothesis space.
	No, the answer is incorrect. Score: 0
	Accepted Answers: B) Atleast one set of 4 points can be shattered by the hypothesis space. D) No set of 5 points can be shattered by the hypothesis space.
	7) Computational complexity of classes of learning problems depends on which of the following?
	A) The size or complexity of the hypothesis space considered by learner
	B) the accuracy to which the target concept must be approximated
	C) the probability that the learner will output a successful hypothesis

D) All of the above	
No, the answer is incorrect. Score: 0	
Accepted Answers:	
D) All of the above	0 4
8) Consider a circle in 2D whose center is at the origin. What is its VC dimension	? 1 point
A) 1	
В) 2	
C) 3	
O) 4	
No, the answer is incorrect. Score: 0	
Accepted Answers: B) 2	
9) Which among the following prevents overfitting when we perform bagging?	1 point
A) The use of sampling with replacement as the sampling technique	
B) The use of weak classifiers	
C) The use of classification algorithms which are not prone to overfitting	
D) The practice of validation performed on every classifier trained	
No, the answer is incorrect. Score: 0	
Accepted Answers: A) The use of sampling with replacement as the sampling technique	
10)/C dimension for conjunctions of n Boolean literals is:	1 point
A) Atleast n	
B) Atmost n	
C) Can't say	
D) None	
No, the answer is incorrect. Score: 0	
Accepted Answers: D) None	
Previous Page	End

X

NIPTEIL

reviewer3@nptel.iitm.ac.in ▼

Courses » Introduction to Machine Learning

Announcements Course Ask a Question Progress Mentor FAQ

Unit 9 - Week 8

Course outline	Assignment 8
	The due date for submitting this assignment has passed.
low to access he portal ?	As per our records you have not submitted this assignment. Due on 2018-10-24, 23:59 IST.
Veek 1	1) Suppose you run K-means clustering algorithm on a given dataset. What are the factors on <i>1 poin</i> which the final clusters depend on ?
Veek 2	I. The value of K II. The initial cluster seeds chosen
Week 3	III. The distance function used.
Veek 4	A) I only
Week 5	B) II only
Veek 6	C) I and II only
Veek 7	D) I, II and III
	No, the answer is incorrect. Score: 0
Veek 8	
Lecture 41:	Accepted Answers: D) I, II and III
Introduction to Clustering	2) Which of the following statements are true about the different types of linkages. 1 <i>poi</i> .
Lecture 42: Kmeans	A. single linkage suffers from chaining.
Clustering	B. Average linkage suffers from crowding.
Lecture 43: Agglomerative Hierarchical	C. In single linkage clustering the similarity between two clusters depends on all the element in the two clusters.
Clustering	D. Complete linkage avoids chaining but suffers from crowding.
Lecture 44:	No, the answer is incorrect.
Exereise on	Score: 0
Kmeans	Accepted Answers:
Clustering	A. single linkage suffers from chaining.

© 2014 NPTEL - Privacy & Terms - Honor Code - FAQs -

NPTEL National Programme on Technology Enhanced Learning



Funded by

In association with

A project of

Videos		2. Only in single-link clustering	
Assignment	ce De	3. Only in complete-link clustering	
Solution		4. In both single-link and complete-link clustering	
		No, the answer is incorrect.	
		Score: 0	
		Accepted Answers:	
		4. In both single-link and complete-link clustering	
		Choose ALL the statements that are true for hierarchical agglomerative clustering	0 points
		A) The number of clusters need to be pre-specified.	
		B) The output of the clustering algorithm depends oil the chclce.of the similarity metr	ic.
		C) The number of merge operations depends on the number of clusters desired.	
		D) The number of merge operations depends on the characteristics of the data set.	
		No, the answer is incorrect. Score: 0	
		Accepted Answers:	
		B) The output of the clustering algorithm depends oil the chclce.of the similarity metric.C) The number of merge operations depends on the number of clusters desired.	
		5)	1 point
		We would like to cluster the natural numbers from 1 to 1024 into two clusters using hierarchical agglomerative clustering. We will use Euclidian distance as our distance measure. We break ties by merging the clusters in which the lowest natural number resides. For example, if the distance between clusters A and B is the same as the distance between cluster C and D, we would choose A and B as the next clusters to merge if	
		$\min[A,B] < \min[C,D]$, where {A,B} are the set of natural numbers assigned to clusters A and B. For complete linkage clustering method specify the number of elements assigned to each of the clusters obtained by cutting the dendrogram at the root. In complete linkage clustering the distance between two clusters is the distance of the farthest members of the clusters.	
		A) 1,1023	
		B) 512,512	
		C) 1022,2	
		D) None of these	
		No, the answer is incorrect.	
		Score: 0	
		Accepted Answers:	
		B) 512,512	
		6) Which of the following is riot a clustering approach	0 points
		A) Partitioning	
		B) Hierarchical	
		C) Density-based	
		D) Bagging	
		No, the answer is incorrect.	
		Score: 0	
		Accepted Answers: D) Bagging	
		7) Which of the following options is a measure of internal evaluation of a clustering algorithm	? 1 point

A) Rand index	
B) Davies-Bouldin index	
C) Jaccard index	
D) F-measure	
No, the answer is incorrect. Score: 0	
Accepted Answers:	
B) Davies-Bouldin index	
8) Which among the following is/are some of the assumptions made by the k-means algorithm (assuming Euclidean distance measure)?	oints
A) Clusters are spherical in shape	
B) Clusters are of similar sizes	
C) Data points in one cluster are well separated from data points of other clusters	
D) There is no wide variation in density among the data points	
No, the answer is incorrect. Score: 0	
Accepted Answers:	
A) Clusters are spherical in shape B) Clusters are of similar sizes	
	point
We are given the following four data points in two dimension: x_1 = (2,2), x_2 =(8,6), x_3 =(6,8), x_4 = (2,4). We want to cluster the data points into two clusters C_1 and C_2 using the K-Means algorithm. Euclidean distance is used for clustering. To initialize the algorithm we consider C_1 ={ x_1, x_3 } and C_2 ={ x_2, x_4 }. After two iteration of the K-means algorithm, the cluster memberships are:	
a. $C_1 = \{x_1, x_2\} \text{ and } C_2 = \{x_3, x_4\}$	
$C_1 = \{x_1, x_4\} \text{ and } C_2 = \{x_2, x_3\}$	
$C_1 = \{x_1, x_3\} \text{ and } C_2 = \{x_2, x_4\}$	
d. None of these.	
No, the answer is incorrect. Score: 0	
Accepted Answers:	
$C_1 = \{x_1, x_4\} \text{ and } C_2 = \{x_2, x_3\}$	
10)With respect to k-means clustering, which of the following are the correct descriptions of the 1 expectation (E) and maximization (M) steps respectively?	point
 A) E-step: assign points to nearest cluster center, M-step: estimate model parameters the maximize the likelihood for the given assignment of points. 	ıat
B) E-step: estimate model parameters that maxim, ize the likelihood for the given assign of points, M-step: assign points to nearest clustercenter.	iment
C) None of A or B.	
D) Both A and B	
No, the answer is incorrect. Score: 0	
Accepted Answers: A) E-step: assign points to nearest cluster center, M-step: estimate model parameters that maximum.	imize th

Previous Page

End