

Unit 2 - Week 1

Course outline

How to access the portal ?

Week 1

- ☐ Lecture 01: Introduction
- ☐ Lecture 02 : Different Types of Learning
- ☐ Lecture 03 : Hypothesis Space and Inductive Bias
- ☐ Lecture 04 : Evaluation and Cross-Validation
- ☐ Lecture 5: Tutorial - I
- ☐ Quiz : Week 1 Assignment 1
- ☐ Week 1: Assignment Solution
- ☐ Feedback for Week 1

Week 2

Week 3:

Week 4

Week 5

Week 6

Week 7

Week 8:

Week 1 Assignment 1

The due date for submitting this assignment has passed.

Due on 2017-08-07, 23:59 IST.

Assignment submitted on 2017-08-01, 01:19 IST

1) Which ONE of the following are regression tasks?

- ☒ A) Predict the age of a person
- ☐ B) Predict the country from where the person comes from
- ☐ C) Predict whether the price of petroleum will increase tomorrow
- ☐ D) Predict whether a document is related to science

Yes, the answer is correct.

Score: 1

Accepted Answers:

A) Predict the age of a person

2) Which of the following are classification tasks? (Mark all that apply)

- ☒ A) Find the gender of a person by analyzing his writing style
- ☐ B) Predict the price of a house based on floor area, number of rooms etc.
- ☒ C) Predict whether there will be abnormally heavy rainfall next year
- ☐ D) Predict the number of copies of a book that will be sold this month

Yes, the answer is correct.

Score: 1

Accepted Answers:

A) Find the gender of a person by analyzing his writing style

C) Predict whether there will be abnormally heavy rainfall next year

3) Which of the following are examples of unsupervised learning?

- ☒ A) Group news articles based on text similarity
- ☒ B) Make clusters of books on similar topics in a library
- ☐ C) Filter out spam emails
- ☐ D) Segment online customers into two classes based on their age group – below 25 or above 25

Yes, the answer is correct.

Score: 1

Accepted Answers:

A) Group news articles based on text similarity

B) Make clusters of books on similar topics in a library

4) Which of these are categorical features?

- ☐ A) Height of a person
- ☐ B) Price of petroleum
- ☒ C) Mother tongue of a person
- ☐ D) Amount of rainfall in a day

Yes, the answer is correct.

Score: 1

Accepted Answers:

C) Mother tongue of a person

5) Validation set is used for testing the generalization performance of a learning algorithm.

- ☐ A) T
- ☒ B) F

No, the answer is incorrect.

Score: 0

Accepted Answers:

A) T

6) Which one of the following functions has the highest bias?

2 points

- ☒ A) Linear model
- ☐ B) Quadratic model
- ☐ C) Decision tree

Yes, the answer is correct.

Score: 2

Accepted Answers:

A) Linear model

7) 3. The variance in an existing model can be reduced by _____ training data. (Choose the appropriate option for filling the blank)

2 points

- ☐ A) Decreasing
- ☒ B) Increasing

Yes, the answer is correct.

Score: 2

Accepted Answers:

B) Increasing

8) We always wish to make sure that our model performs well in the real world when presented with cases that it has not encountered before. Hence we make sure the _____ is low. (Choose the appropriate option for filling the blank):

2 points

- ☒ A) Bias
- ☐ B) Variance

No, the answer is incorrect.

Score: 0

Accepted Answers:

B) Variance

9) 1. I am the marketing consultant of a leading e-commerce website. I have been given a task of making a system that recommends products to users based on their activity on Facebook. I realize that user-interests could be highly variable. Hence I decide to a. First, cluster the users into communities of like-minded people and b. Second, train separate models for each community to predict which product category (e.g. electronic gadgets, cosmetics, etc) would be the most relevant to that community. The first task is a/an _____ learning problem while the second is a/an _____ problem. Choose from the options:

2 points

- ☐ A) Supervised and unsupervised
- ☒ B) Unsupervised and supervised
- ☐ C) Supervised and supervised
- ☐ D) Unsupervised and unsupervised

Yes, the answer is correct.

Score: 2

Accepted Answers:

B) Unsupervised and supervised

10) At what value of number of nodes does overfitting set in?

2 points

Choose a number from among the following that is the closest to what you expect:

- ☐ A) 5
- ☒ B) 20
- ☐ C) 30

Yes, the answer is correct.

Score: 2

Accepted Answers:

B) 20

11) Choose the function that has the maximum variance

2 points

- ☒ A)
- ☐ B)
- ☐ C)

No, the answer is incorrect.

Score: 0

Accepted Answers:

C)

[Previous Page](#)

[End](#)

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

A project of



NPTEL

National Programme on
Technology Enhanced Learning

In association with

NASSCOM®

Powered by

Google™

Funded by

Government of India
Ministry of Human Resource Development

Unit 3 - Week 2

Course outline

How to access the portal ?

Week 1

Week 2

- ☐ Lecture 06 : Linear Regression
- ☐ Lecture 07 : Introduction to Decision Trees
- ☐ Lecture 08 : Learning Decision Tree
- ☐ Lecture 09 : Overfitting
- ☐ Lecture 10: Python Exercise on Decision Tree and Linear Regression
- ☐ Lecture 11: Tutorial - II
- ☐ Quiz : Week 2 Assignment 1
- ☐ Lecture notes - Week 2
- ☐ Week 2: Assignment Solution
- ☐ Feedback for Week 2

Week 3:

Week 4

Week 5

Week 6

Week 7

Week 8:

Week 2 Assignment 1

The due date for submitting this assignment has passed.
As per our records you have not submitted this assignment.

Due on 2017-08-11, 23:59 IST.

1) 1. In regression the output is

- ☐ A) Discrete.
- ☐ B) Continuous and always lies in a finite range.
- ☐ C) Continuous.
- ☐ D) May be discrete or continuous.

No, the answer is incorrect.

Score: 0

Accepted Answers:

C) Continuous.

2) In linear regression the parameters are

- ☐ A) strictly integers
- ☐ B) always lies in the range [0,1]
- ☐ C) any value in the real space
- ☐ D) any value in the complex space

No, the answer is incorrect.

Score: 0

Accepted Answers:

C) any value in the real space

3) Which of the following is true for a decision tree?

- ☐ A) Decision tree is an example of linear classifier.
- ☐ B) The entropy of a node typically decreases as we go down a decision tree.
- ☐ C) Entropy is a measure of purity.
- ☐ D) An attribute with lower mutual information should be preferred to other attributes.

No, the answer is incorrect.

Score: 0

Accepted Answers:

B) The entropy of a node typically decreases as we go down a decision tree.

4) 2. Given $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, best fitting data to $y = f(x)$ by least squares requires minimization of

- ☐ ☐
- ☐ ☐
- ☐ ☐
- ☐ ☐

No, the answer is incorrect.

Score: 0

Accepted Answers:

☐

5) Given a list of 14 examples including 9 positive and 5 negative examples. The entropy of the dataset with respect to this classification is

- ☐ A) 0.940
- ☐ B) 0.06
- ☐ C) 0.50
- ☐ D) 0.22

No, the answer is incorrect.

Score: 0

Accepted Answers:

A) 0.940

6) What is the value of information gain in the following partitioning?

- ☐ A) 0.72
- ☐ B) 0.42

1 point

1 point

1 point

1 point

2 points

2 points

- ☐ C) 0.28
☐ D) 0.30

No, the answer is incorrect.

Score: 0

Accepted Answers:

C) 0.28

7) The following table shows the results of a recently conducted study on the correlation of the number of hours spent driving with the risk of developing acute back-ache. Find the equation of the best fit line for this data. **2 points**

- ☐ 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

Accepted Answers:

C) $y = 4.59x + 12.58$

Programming question:

A dataset collected in a cosmetics shop showing details of customers and whether or not they responded to a special offer to buy a new lip-stick is shown in table below. Use this dataset to build a decision tree, with Buys as the target variable, to help in buying lip-sticks in the future.

You can use [sklearn.tree.DecisionTreeClassifier](#) for solving the problem.

Please download the data for this question

<https://drive.google.com/file/d/0B1wemHAii3JQcVBQTnNaXzZyd1U/view?usp=sharing> (the qualitative fields of the data in the table above has been converted into numbers) and place it in your present working directory. The following python code will load the inputs and targets from the .txt file into the numpy matrices x_train and y_train:

```
import numpy as np
f = open('decision_tree_data.txt','r')
x_train = []
y_train = []

for line in f:
    line = np.asarray(line.split(),dtype=np.float32)
    x_train.append(line[:-1])
    y_train.append(line[-1])

x_train = np.asmatrix(x_train)
y_train = np.reshape(y_train,(len(y_train),1))
```

Now answer the questions 9 and 10 about the decision tree you made:

8) According to the decision tree you have made from previous training data set, what is the decision for the test data: **2 points**

[Age < 21, Income = Low, Gender = Female, Marital Status = Married]?

- ☐ A) Yes
☐ B) No

No, the answer is incorrect.

Score: 0

Accepted Answers:

A) Yes

Previous Page

End

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

A project of



NPTEL

National Programme on
Technology Enhanced Learning

In association with

NASSCOM[®]

Funded by

Government of India
Ministry of Human Resource Development

Powered by

Google[™]

Unit 4 - Week 3:

Course outline

How to access the portal ?

Week 1

Week 2

Week 3:

- ☐ Lecture 12: k-Nearest Neighbour
- ☐ Lecture 13: Feature Selection
- ☐ Lecture 14: Feature Extraction
- ☐ Lecture 15: Collaborative Filtering
- ☐ Lecture 16: Python Exercise on kNN and PCA
- ☐ Lecture 17: Tutorial III
- ☐ Quiz : Week 3 Assignment 1
- ☐ Lecture notes - Week 3
- ☐ Week 3 Assignment 1 Solution
- ☐ Feedback for Week 3

Week 4

Week 5

Week 6

Week 7

Week 8:

Week 3 Assignment 1

The due date for submitting this assignment has passed.
As per our records you have not submitted this assignment.

Due on 2017-08-16, 23:59 IST.

- 1) In k -NN algorithm, given a set of training examples and the value of $k < \text{size of training set}$ (n), the algorithm predicts the class of a test example to be the **1 point**
- ☐ A) most frequent class among the classes of k closest training examples.
 - ☐ B) least frequent class among the classes of k closest training examples.
 - ☐ C) class of the closest point.
 - ☐ D) most frequent class among the classes of the k farthest training examples.

No, the answer is incorrect.

Score: 0

Accepted Answers:

A) most frequent class among the classes of k closest training examples.

- 2) Which of the following is NOT TRUE for memory-based learning? **1 point**
- ☐ A) It learns a global model
 - ☐ B) Cross-validation is very efficient
 - ☐ C) The runtime cost scales with training data
 - ☐ D) Training is fast

No, the answer is incorrect.

Score: 0

Accepted Answers:

A) It learns a global model

- 3) For which of the following cases feature selection may be used? **1 point**
- ☐ A) If large number of completely irrelevant features are present in the data.
 - ☐ B) For better interpretability of data.
 - ☐ 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

- 4) Given N features, where $N=200$, it is possible to find the best subset of features under practical situations. **1 point**
- ☐ A) True
 - ☐ B) False

No, the answer is incorrect.

Score: 0

Accepted Answers:

B) False

- 5) Consider the figures below. Which figure shows the most probable PCA component directions for the data points? **1 point**

- ☐ A) A
- ☐ B) B
- ☐ C) C
- ☐ D) D

No, the answer is incorrect.

Score: 0

Accepted Answers:

A) A

6) Imagine that the centre of the target is a model that perfectly predicts the correct values. As we move away from the bulls-eye, our predictions get worse and worse. Two different distributions of training data points are given in the two diagrams below. What can you say about the bias and variance of the final model trained with these training sets?

- ☐ A) I – High bias, Low variance, II – High bias, high variance
☐ B) I – Low bias, Low variance, II – High bias, low variance
☐ C) I – Low bias, Low variance, II – High bias, high variance
☐ D) I – High bias, Low variance, II – Low bias, high variance

No, the answer is incorrect.

Score: 0

Accepted Answers:

C) I – Low bias, Low variance, II – High bias, high variance

7) 8. In the following diagram the circles and the squares represent instances of class I and class II respectively. Among the blue lines which one is the most probable LDA direction?

- ☐ A) L1
☐ B) L2
☐ C) L3
☐ D) L4

No, the answer is incorrect.

Score: 0

Accepted Answers:

A) L1

8) In the following diagram let blue circles indicate positive examples and orange squares indicate negative examples. We want to use kNN algorithm for classifying the points. If $k=3$, find the class of the point (6,6).

- ☐ A) Positive
☐ B) Negative

No, the answer is incorrect.

Score: 0

Accepted Answers:

B) Negative

The following table shows the ratings given by viewers A, B and C to the movies HP1, HP2, HP3, TW, SW1, SW2 and SW3 in the scale of 1-5. Blank entry implies that the viewer has not watched the movie. Based on this data answer Questions 9, 10 and 11.

	HP1	HP2	HP3	TW	SW1	SW2	SW3
A	4			5	1		
B	5	5	4				
C				2	4	5	
D		3					3

9) Using cosine similarity, find the similarity in opinion of A and B

- ☐ A) 0.21
☐ B) 0.1
☐ C) 0.38
☐ D) -0.21

No, the answer is incorrect.

Score: 0

Accepted Answers:

C) 0.38

10) Normalize the table in Question 9 by subtracting from each rating average rating of that user such that low ratings are converted to negative numbers and high ratings are converted to positive numbers. What are the similarity values between A and B, and, A and C after normalization

- ☐ A) 0.092, -0.559
☐ B) -0.559, 0.092
☐ C) 0.521, 0.533
☐ D) 0.533, -0.521

No, the answer is incorrect.

Score: 0

Accepted Answers:

A) 0.092, -0.559

11) Which of the following is advantage of normalization?

1 point

- ☐ A) It increases the distance between the users with different opinion
- ☐ B) It helps to identify the users with useless opinion
- ☐ C) Both A and B
- ☐ D) None of A and B

No, the answer is incorrect.

Score: 0

Accepted Answers:

C) Both A and B

Previous Page

End

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

A project of



NPTEL

National Programme on
Technology Enhanced Learning

In association with

NASSCOM®

Powered by

Google™

Funded by

Government of India
Ministry of Human Resource Development

Unit 5 - Week 4

Course outline

How to access the portal ?

Week 1

Week 2

Week 3:

Week 4

- ☐ Lecture 18: Bayesian Learning
- ☐ Lecture 19: Naive Bayes
- ☐ Lecture 20: Bayesian Network
- ☐ Lecture 21: Python Exercise on Naive Bayes
- ☐ Lecture 22: Tutorial 4
- ☐ Quiz : Week 4 Assignment 1
- ☐ Lecture notes - Week 4
- ☐ Feedback for Week 4
- ☐ Week 4: Assignment 1 Solution

Week 5

Week 6

Week 7

Week 8:

Week 4 Assignment 1

The due date for submitting this assignment has passed.

Due on 2017-08-25, 23:59 IST.

Assignment submitted on 2017-08-25, 21:16 IST

Week 4 Assignment 1 Group 1

1) Which of the following properties is false in the case of a Bayesian Network:

2 points

- ☐ A) The edges are directed
- ☒ B) Contains cycles
- ☐ C) Represents conditional independence relations among random variables
- ☐ D) All of the above

Yes, the answer is correct.

Score: 2

Accepted Answers:

B) Contains cycles

2) A and B are Boolean random variables. Given: $P(A=True) = 0.3$, $P(A=False) = 0.7$, $P(B=True|A=True) = 0.4$, $P(B=False|A=True) = 0.6$, $P(B=True|A=False) = 0.6$, $P(B=False|A=False) = 0.4$. Calculate $P(A=True|B=False)$ by Bayes rule.

- ☐ A) 0.49
- ☒ B) 0.39
- ☐ C) 0.37
- ☐ D) 0.28

Yes, the answer is correct.

Score: 2

Accepted Answers:

B) 0.39

13) In the following Bayesian network A, B and C are Boolean random variables taking values in {True, False}.

2 points

Which of the following statements is true?

- ☐ A) The value of C is not given. If the value of B changes from True to False, the conditional probability of A, $P(A|B)$ changes.
- ☒ B) The value of C is given to be True. If the value of B changes from True to False, the conditional probability of A, $P(A|B)$ changes.
- ☐ C) Neither A nor B
- ☐ D) Both A and B

Yes, the answer is correct.

Score: 2

Accepted Answers:

B) The value of C is given to be True. If the value of B changes from True to False, the conditional probability of A, $P(A|B)$ changes.

4) Diabetic Retinopathy is a disease that affects 80% people who have diabetes for more than 10 years. 5% 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?

2 points

- ☐ A) 0.024
- ☒ B) 0.040
- ☐ C) 0.076
- ☐ D) 0.005

Yes, the answer is correct.

Score: 2

Accepted Answers:

B) 0.040

5) 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 before A throws 6. If A begins, his chance of winning would be:

2 points

- ☒ A) 30/61
- ☐ B) 31/61
- ☐ C) 1/2
- ☐ D) 6/7

Yes, the answer is correct.

Score: 2

Accepted Answers:

A) 30/61

6) Using following dataset answer the next question: □

2 points

Using Naïve Bayes classifier what will be the class when a new species holds following feature values:

- ☒ A) mammals
- ☐ B) non-mammals

Yes, the answer is correct.

Score: 2

Accepted Answers:

A) mammals

7) A problem in mathematics is given to three students Ram, Rahim and Nusrat whose chances of solving are 1/3, 1/4 and 1/2. The probability that the problem will be solved:

2 points

- ☐ A) 1/4
- ☐ B) 3/4
- ☐ C) 4/5
- ☐ D) 1/24

No, the answer is incorrect.

Score: 0

Accepted Answers:

B) 3/4

Previous Page

End

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

A project of



NPTEL

National Programme on
Technology Enhanced Learning

In association with

NASSCOM®

Powered by

Google™

Funded by

Government of India
Ministry of Human Resource Development

Unit 6 - Week 5

Course outline

How to access the portal ?

Week 1

Week 2

Week 3:

Week 4

Week 5

☐ Lecture 23 :
Logistic Regression

☐ Lecture 24:
Introduction
Support Vector
Machine

☐ Lecture 25: SVM :
The Dual
Formulation

☐ Lecture 26: SVM :
Maximum Margin
with Noise

☐ Lecture 27:
Nonlinear SVM and
Kernel Function

☐ Lecture 28: SVM :
Solution to the Dual
Problem

☐ Lecture 29: Python
Exercise on SVM

☐ Lecture notes -
Week 5

☐ Quiz : Week 5
Assignment 1

☐ Feedback for Week
5

☐ Week 5
Assignment
Solution

Week 6

Week 7

Week 8:

Week 5 Assignment 1

The due date for submitting this assignment has passed.
As per our records you have not submitted this assignment.

Due on 2017-09-03, 23:59 IST.

1) Which of the following is true about functional margin of SVM? I. Functional margin of a point (x_i) is measured by its distance from the decision boundary (w, b). II. Larger functional margin implies more confidence for correct prediction. **1 point**

- ☐ A. I only
☐ B. II only
☐ C. None
☐ D. Both I and II

No, the answer is incorrect.

Score: 0

Accepted Answers:

D. Both I and II

2) The dual of the large margin linear classifier problem allow us to use kernels to get optimal margin classifiers to work efficiently in very high dimensional spaces. **1 point**

- ☐ A. True
☐ B. False

No, the answer is incorrect.

Score: 0

Accepted Answers:

A. True

3) Which of the following is NOT a kernel function? **2 points**

- ☐ A. $K(x_i, x_j) = x_i \cdot x_j$
☐ 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. $K(x_i, x_j) = \tanh(\beta_0 x_i \cdot x_j + \beta_1)$

No, the answer is incorrect.

Score: 0

Accepted Answers:

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

4) Which of the following ARE TRUE about SMO algorithm (multiple answers). **2 points**

- ☐ 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.

5) Which among the following is/are the most appropriate kernels that can be used with SVM to separate the classes below. **2 points**

- ☐ A. Linear kernel

- ☐ B. Gaussian RBF kernel
- ☐ C. Both A and B
- ☐ D. None of the above

No, the answer is incorrect.

Score: 0

Accepted Answers:

B. Gaussian RBF kernel

6) The optimization problem for large margin linear binary classifier where w , x and $y \in \{-1, 1\}$ represent the parameter vector, data point and label respectively, is given by, 1 point

- ☐
- A. minimize $\frac{1}{2} \|w\|^2$ such that for all data points $y_i(w^T x_i + b) \geq 1$
- ☐
- B. minimize $\frac{1}{2} \|w\|$ such that for all data points $y_i(w^T x_i + b) \leq 1$
- ☐
- C. minimize $\frac{1}{2} \|w\|^2$ such that for all data points $y_i(w^T x_i + b) \leq 1$
- ☐
- D. maximize $\frac{1}{2} \|w\|$ such that for all data points $y_i(w^T x_i + b) \geq 1$

No, the answer is incorrect.

Score: 0

Accepted Answers:

A. minimize $\frac{1}{2} \|w\|^2$ such that for all data points $y_i(w^T x_i + b) \geq 1$

7) Which kernels can we use in SVM to separate the two classes (denoted by '+' and '-') in the figure below. 2 points

- ☐ A. Linear kernel
- ☐ B. Quadratic kernel
- ☐ C. Both A and B
- ☐ D. None

No, the answer is incorrect.

Score: 0

Accepted Answers:

B. Quadratic kernel

8) The diagram below shows data points from two different classes (denoted by '+' and '-'). If an SVM is trained on the data, which points will act as the support vectors? 2 points

- ☐
- A. x_1 and x_2
- ☐
- B. x_2 and x_3
- ☐
- C. x_3 and x_4
- ☐
- D. x_1 and x_4

No, the answer is incorrect.

Score: 0

Accepted Answers:

B. x_2 and x_3

9) Consider building an SVM over the dataset shown in the figure below. 2 points

The equation of the hyperplane is given by

- ☐ A. $w = (4/5, 2/5)$, $b = -11/5$
- ☐ B. $w = (2/5, 4/5)$, $b = -11/5$
- ☐ C. $w = (-2/5, -4/5)$, $b = -11/5$
- ☐ D. $w = (2/5, 4/5)$, $b = 11/5$

No, the answer is incorrect.

Score: 0

Accepted Answers:

B. $w = (2/5, 4/5)$, $b = -11/5$

The data for programming question 10 and 11 may be downloaded from this link (<https://drive.google.com/file/d/0B1wemHAii3JQR25kR2Z5UTVEOVU/view?usp=sharing>).

Load the data using the code given in the link (<https://drive.google.com/file/d/0B1wemHAii3JQeEEwZ25hRld2bW8/view?usp=sharing>). Train SVM using RBF kernel, linear kernel and polynomial kernel of degree 3 using scikit-learn.
Set the value of parameter C to 10.0. Based on above information answer question 10 and 11.

10) How many support vectors obtained using rbf, linear and polynomial of degree 3 kernels respectively on the given data? **2 points**

- ☐ A. 3,81,3
- ☐ B. 81,3,81
- ☐ C. 81,3,3
- ☐ D. 3,3,3

No, the answer is incorrect.

Score: 0

Accepted Answers:

C. 81,3,3

11) Which of the following point is common among the support vectors obtained using linear kernel and polynomial kernel of degree 3? **2 points**

- ☐ A. [19.96749215, 11.77640679]
- ☐ B. [2.34752767, 4.5295017]
- ☐ C. [12.11631344, 17.70352844]
- ☐ D. None of the above

No, the answer is incorrect.

Score: 0

Accepted Answers:

C. [12.11631344, 17.70352844]

Previous Page

End

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

A project of



NPTEL

National Programme on
Technology Enhanced Learning

In association with

NASSCOM®

Powered by

Google™

Funded by

Government of India
Ministry of Human Resource Development

Unit 7 - Week 6

Course outline

How to access the portal ?

Week 1

Week 2

Week 3:

Week 4

Week 5

Week 6

- ☐ Lecture 30 : Introduction
- ☐ Lecture 31 : Multilayer Neural Network
- ☐ Lecture 32: Neural Network and Backpropagation Algorithm
- ☐ Lecture 33: Deep Neural Network
- ☐ Lecture 34 : Python Exercise on Neural Network
- ☐ Lecture 35: Tutorial 6
- ☐ Lecture notes - Week 6
- ☐ Quiz : Week 6: Assignment 1
- ☐ Feedback for Week 6
- ☐ Week 6: Assignment Solution

Week 7

Week 8:

Week 6: Assignment 1

The due date for submitting this assignment has passed.
As per our records you have not submitted this assignment.

Due on 2017-09-09, 23:59 IST.

1) A 4-input neuron has bias of 0 and weights 1, 2, 3 and 4. The transfer function is given by $f(v) = \max(0, v)$. The inputs are 4, 10, 5 and 20 respectively. The output will be **1 point**

- ☐ A) 238
- ☐ B) 76
- ☐ C) 119
- ☐ D) 121

No, the answer is incorrect.

Score: 0

Accepted Answers:

C) 119

2) The back-propagation learning algorithm applied to a two layer neural network **1 point**

- ☐ A) always finds the globally optimal solution.
- ☐ B) finds a locally optimal solution which may be globally optimal.
- ☐ C) never finds the globally optimal solution.
- ☐ D) finds a locally optimal solution which is never globally optimal

No, the answer is incorrect.

Score: 0

Accepted Answers:

B) finds a locally optimal solution which may be globally optimal.

13) The configuration of points belonging to two classes as shown in the diagram below can be shattered by **1 point**

- ☐ A) Linear kernel SVM
- ☐ B) A perceptron
- ☐ C) A single hidden layer neural network with two hidden nodes
- ☐ D) All of the above

No, the answer is incorrect.

Score: 0

Accepted Answers:

C) A single hidden layer neural network with two hidden nodes

4) 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.

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.

5) Which of the following statements is false: **1 point**

- ☐ A) The chances of overfitting decrease with Increasing the number of hidden nodes and increasing the number of hidden layers.

- ☐ B) A neural network with one hidden layer can represent any Boolean function given sufficient number of hidden units and appropriate activation functions.
- ☐ C) Two layer neural networks can represent any continuous functions (within a tolerance) as long as the number of hidden units is sufficient and appropriate activation functions used.

No, the answer is incorrect.

Score: 0

Accepted Answers:

A) The chances of overfitting decrease with increasing the number of hidden nodes and increasing the number of hidden layers.

6) What would be a correct option for weight W [w_0, w_1, w_2], so that the following sigmoid unit will function as an AND gate? 2 points

- ☐ A) [-5,10,10]
- ☐ B) [-10,15,15]
- ☐ C) [-10,15,5]
- ☐ D) [-10,9,9]

No, the answer is incorrect.

Score: 0

Accepted Answers:

D) [-10,9,9]

7) Which one of the following sets of values for the bias (b) would enable the perceptron in the figure below to realize the Boolean function described in the following truth table? Assume all activation functions to be the threshold function which is 1 for all input values greater than zero and 0, otherwise. 2 points

- ☐ A) 0
- ☐ B) 1
- ☐ C) -1.7
- ☐ D) -2.2

No, the answer is incorrect.

Score: 0

Accepted Answers:

C) -1.7

8) The figure shows a single hidden layer neural network. The weights are initialized to 1's as shown in the diagram and all biases are initialized to 0's. Assume all the neurons have linear activation functions. The neural network is to be trained with stochastic (online) gradient descent. The first training example is [$x_1=1, x_2=0$] and the desired output is 1. Calculate the updated value for W_{11} after backpropagation. Choose the value that is the closest to the options given below: [learning rate = 0.1] 5 points

- ☐ A) 0.9
- ☐ B) 1.1
- ☐ C) 1
- ☐ D) -0.1

No, the answer is incorrect.

Score: 0

Accepted Answers:

A) 0.9

Previous Page

End

A project of



NPTEL

National Programme on
Technology Enhanced Learning

In association with

NASSCOM[®]

Powered by

Funded by

Government of India
Ministry of Human Resource Development

Google[™]

Unit 8 - Week 7

Course outline

How to access the portal ?

Week 1

Week 2

Week 3:

Week 4

Week 5

Week 6

Week 7

- ☐ Lecture 36 : Introduction to Computational Learning Theory
- ☐ Lecture 37 : Sample Complexity : Finite Hypothesis Space
- ☐ Lecture 38: VC Dimension
- ☐ Lecture 39: Introduction to Ensembles
- ☐ Lecture 40: Bagging and Boosting
- ☐ Tutorial 7
- ☐ Lecture Notes - Week 7
- ☐ Quiz : Week 7 Assignment 1
- ☐ Week 7 Lecture Notes
- ☐ Week 7: Assignment Slution

Week 8:

Week 7 Assignment 1

The due date for submitting this assignment has passed.
As per our records you have not submitted this assignment.

Due on 2017-09-16, 23:59 IST.

1) The VC dimension of hypothesis space H_1 is larger than the VC dimension of hypothesis space H_2 . Which of the following can be inferred from this? **1 point**

- ☐ The number of examples required for learning a hypothesis in H_1 is larger than the number of examples required for H_2 .
- ☐ The number of examples required for learning a hypothesis in H_1 is smaller than the number of examples required for H_2 .
- ☐ No relation to number of samples required for PAC learning.

No, the answer is incorrect.

Score: 0

Accepted Answers:

The number of examples required for learning a hypothesis in H_1 is larger than the number of examples required for H_2 .

2) The Bayes Optimal Classifier **1 point**

- ☐ is an ensemble of some selected hypotheses in the hypothesis space.
- ☐ is an ensemble of all the hypotheses in the hypothesis space.
- ☐ is the hypothesis that gives best result on test instances.
- ☐ none of the above

No, the answer is incorrect.

Score: 0

Accepted Answers:

is an ensemble of all the hypotheses in the hypothesis space.

3) For a particular learning task, if the requirement of error parameter ϵ changes from 0.1 to 0.01. How many more samples will be required for PAC learning? **1 point**

- ☐ Same
- ☐ 2 times
- ☐ 10 times
- ☐ 100 times

No, the answer is incorrect.

Score: 0

Accepted Answers:

10 times

4) Suppose the VC dimension of a hypothesis space is 4. Which of the following are true? **2 points**

- ☐ No sets of 4 points can be shattered by the hypothesis space.
- ☐ Atleast one set of 4 points can be shattered by the hypothesis space.
- ☐ All sets of 4 points can be shattered by the hypothesis space.
- ☐ No set of 5 points can be shattered by the hypothesis space.

No, the answer is incorrect.

Score: 0

Accepted Answers:

Atleast one set of 4 points can be shattered by the hypothesis space.

No set of 5 points can be shattered by the hypothesis space.

5) Consider a circle in 2D whose center is at the origin. What is its VC dimension? **2 points**

- ☐ 1
- ☐ 2
- ☐ 3
- ☐ 4

No, the answer is incorrect.

Score: 0

Accepted Answers:

2

6) Under a binary classification setting, which of the following sets of three labeled points cannot be shattered by a circle centered at the origin?

2 points

- ☐ A
- ☐ B
- ☐ C
- ☐ D

No, the answer is incorrect.

Score: 0

Accepted Answers:

C

7) Given a set of 4 points $(x1,-)$, $(x2,-)$, $(x3,+)$ and $(x4,+)$, Adaboost algorithm is used to train a weak classifier on this data. In the first iteration, the weak classifier wrongly classifies $x3$ and correctly classifies the other three points. In the second iteration it wrongly classifies $x1$ and correctly classifies the other points. Assuming uniform initial weight distribution ($D1$) over the data points, what is the weight distribution for the 3rd iteration ($D3$).

2 points

- ☐ 0.5, 0.3, 0.1, 0.1
- ☐ 0.1, 0.3, 0.5, 0.1
- ☐ 0.5, 0.1, 0.3, 0.1
- ☐ 0.1, 0.1, 0.3, 0.5

No, the answer is incorrect.

Score: 0

Accepted Answers:

0.5, 0.1, 0.3, 0.1

Previous Page

End

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

A project of



NPTEL

National Programme on
Technology Enhanced Learning

In association with

NASSCOM®

Powered by

Google™

Funded by

Government of India
Ministry of Human Resource Development

Unit 9 - Week 8:

Course outline

How to access the portal ?

Week 1

Week 2

Week 3:

Week 4

Week 5

Week 6

Week 7

Week 8:

- ☐ Lecture 41: Introduction to Clustering
- ☐ Lecture 42: Kmeans Clustering
- ☐ Lecture 43: Agglomerative Hierarchical Clustering
- ☐ Lecture 44: Python Exercise on Kmeans Clustering
- ☐ Week 8 - Lecture Notes
- ☐ Tutorial 8
- ☐ Quiz : Week 8 : Assignment 1
- ☐ Week 8 : Assignment Solution

Week 8 : Assignment 1

The due date for submitting this assignment has passed.
As per our records you have not submitted this assignment.

Due on 2017-09-20, 23:59 IST.

Week 8 Assignment 1

1) With respect to k-means clustering, which of the following are the correct descriptions of the expectation2 points
(E) and maximization (M) steps respectively?

- ☐ A. E-step: assign points to nearest cluster center, M-step: estimate model parameters that maximize the likelihood for the given assignment of points.
- ☐ B. E-step: estimate model parameters that maximize the likelihood for the given assignment of points, M-step: assign points to nearest cluster center.
- ☐ 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 maximize the likelihood for the given assignment of points.

2) You are given a set of 6 points, {A, B, C, D, E, F} and the distance matrix of size 6 by 6.5 points

This distance matrix was calculated based on the features of the points. Using single linkage clustering construct the dendrogram for the set of points. The final dendrogram is given by which of the following? (The bracing in the options directly corresponds to the dendrogram hierarchical structure.)

- ☐ A. (((D, F), E), B), (A, C))
- ☐ B. (((A, F), E), C), (D, B))
- ☐ C. (((D, F), E), C), (A, B))
- ☐ D. (((D, E), F), C), (A, B))

No, the answer is incorrect.

Score: 0

Accepted Answers:

C. (((D, F), E), C), (A, B))

3) Which of the following options is a measure of internal evaluation of a clustering algorithm?2 points

- ☐ 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

4) K-means clustering is not an example of which of the following clustering method:2 points

- ☐ A. Non-hierarchical clustering
- ☐ B. optimizing partitioning
- ☐ C. Divisive clustering
- ☐ D. Agglomerative clustering

No, the answer is incorrect.

Score: 0

Accepted Answers:

D. Agglomerative clustering

5) Which of the following statements are true about the different types of linkages.2 points

- ☐ A. single linkage suffers from chaining.

- ☐ B. Average linkage suffers from crowding.
- ☐ C. In single linkage clustering the similarity between two clusters depends on all the elements in the two clusters.
- ☐ D. Complete linkage avoids chaining but suffers from crowding.

No, the answer is incorrect.

Score: 0

Accepted Answers:

A. single linkage suffers from chaining.

D. Complete linkage avoids chaining but suffers from crowding.

6) Suppose you run K-means clustering algorithm on a given dataset. What are the factors on which the final clusters depend on ? **3 points**

- I. The value of K
- II. The initial cluster seeds chosen
- III. The distance function used.

- ☐ A. I only
- ☐ B. II only
- ☐ C. I and II only
- ☐ D. I, II and III

No, the answer is incorrect.

Score: 0

Accepted Answers:

D. I, II and III

[Previous Page](#)

[End](#)

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

A project of



NPTEL

National Programme on
Technology Enhanced Learning

In association with

NASSCOM[®]

Powered by

Google[™]

Funded by

Government of India
Ministry of Human Resource Development