

X

reviewer2@nptel.iitm.ac.in ▼

Courses » Introduction to Data Analytics

Announcements

Course

Ask a Question

Progress



# Unit 6 - Week 5 - Supervised Learning (Regression and Classification Techniques) - I

## Course outline

How to access the portal

Week 1 - Course Overview and Descriptive Statistics

Week 2 - Probability Distributions & Inferential Statistics

Week 3 - Inferential Statistics

Week 4 - Machine Learning

Week 5 - Supervised Learning (Regression and Classification Techniques) - I

Logistic Regression

Training a Logistic Regression Classifier

Classification and Regression Trees(Decision Trees)

Classification and Regression Trees(cont'd) - Decision Trees

Bias Variance Dichotomy

Model Assessment and Selection

Support Vector Machines

Support Vector Machines(cont'd)

Support Vector Machines for Non Linearly Separable Data

Support Vector Machines and Kernel Transformations

Tutorial on Decision Trees

Quiz : Assignment 5

Feedback for week 5

Assignment 5: Solution

Week 6 : Supervised Learning (Regression and Classification Techniques)-II

Week 7 - Association Rule Mining and Big Data

Week 8 - Clustering Analysis and Prescriptive Analytics

Course Summary+ Insight into the Final Exam

## Assignment 5

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

Due on 2017-09-01, 23:55 IST

1) In a binary classification scenario where  $x$  is the independent variable and  $y$  is the dependent variable, logistic regression assumes that the conditional distribution  $y|x$  follows a 1 point

- bernoulli distribution
- binomial distribution
- normal distribution
- exponential distribution

No, the answer is incorrect.  
Score: 0

Accepted Answers:  
*bernoulli distribution*

2) What does decision node illustrates in a decision tree? 1 point

- Class of instance
- Data value description
- Test specification
- Data process description

No, the answer is incorrect.  
Score: 0

Accepted Answers:  
*Test specification*

3) Suppose you are given all pages of Wikipedia and your task is to classify each page among one of the  $k$  categories. This classification should be based on the text present on the web-pages. You decided to perform SVM on this data. Which of the following kernel would you prefer? 1 point

- linear kernel
- polynomial kernel
- Gaussian kernel
- all of the above are equally preferable

No, the answer is incorrect.  
Score: 0

Accepted Answers:  
*linear kernel*

4) For a supervised learning problem, we have seen that the data used in the testing phase, i.e., the test set, is not used for training or building the model. Does this mean that we do not require labels for data points in the test set? 1 point

- no
- yes

No, the answer is incorrect.  
Score: 0

Accepted Answers:  
*no*

5) . A 1-NN (KNN with  $K=1$ ) classifier has higher variance than a 3-NN classifier. 1 point

- true
- false

No, the answer is incorrect.  
Score: 0

Accepted Answers:  
*true*

6) In the linearly non-separable case, what effect does the  $C$  parameter have on the SVM model? 1 point

- it determines the count of support vectors
- it is a count of the number of data points which do not lie on their respective side of the hyperplane
- it determines how many data points lie within the margin
- it allows us to trade-off the number of misclassified points in the training data and the size of the margin

No, the answer is incorrect.

Score: 0

Accepted Answers:

it allows us to trade-off the number of misclassified points in the training data and the size of the margin

7) Consider the following data set:

1 point

Day	Outlook	Temperature	Humidity	Wind
D1	Sunny	Hot	High	Weak 
D2	Sunny	Hot	High	Strong 
D3	Overcast	Hot	High	Weak 
D4	Rain	Mild	High	Weak 
D5	Rain	Cool	Normal	Weak 
D6	Rain	Cool	Normal	Strong
D7	Overcast	Cool	Normal	Strong
D8	Sunny	Mild	High	Weak
D9	Sunny	Cool	Normal	Weak
D10	Rain	Mild	Normal	Weak
D11	Sunny	Mild	Normal	Strong
D12	Overcast	Mild	High	Strong
D13	Overcast	Hot	Normal	Weak
D14	Rain	Mild	High	Strong

Considering 'PlayTennis' as the binary values attribute we are trying to predict, which of the attributes would you select as the root in a decision tree with multi-way splits using the information gain measure?

- Humidity
- Wind
- Temperature
- Outlook

No, the answer is incorrect.

Score: 0

Accepted Answers:

Outlook

8) While constructing a decision tree, if we reach at the level where all the leaves are pure leaves, then we stop splitting the leaves further. In the previous example, you chose one attribute as the splitting attribute. After splitting on that attribute, do you need further splitting? [Note: A pure node means, all the data instances in that node have same class] 1 point

- Yes
- No

No, the answer is incorrect.

Score: 0

Accepted Answers:

Yes

**WEKA based questions:** To answer the following questions, you have to use WEKA software.

IRIS data set: You can download IRIS data set from the following link:

[https://osdn.net/projects/sfnet\\_irisds/downloads/IRIS.csv/](https://osdn.net/projects/sfnet_irisds/downloads/IRIS.csv/)

The IRIS data set is a 8 dimensional data set with 9th column as the class of IRIS. There are three possible classes of the IRIS. The data set is in csv format which you can load directly to the WEKA.

**Task 1:** After loading the data, you have to run J48 classifier (decision tree) and answer the following questions.

9) If you train an unpruned decision tree on IRIS data set, what classification accuracy do you get on the training data? 1 point

(Note 1: Set minNumObj = 1)

(Note 2: You can ignore the fractional part in the final answer)

- 33%
- 50%
- 100%
- 92%

No, the answer is incorrect.

Score: 0

Accepted Answers:

100%

**Task 2:** After loading the IRIS data set, you have to learn SVM (SMO in WEKA). Your task is to learn SVM with second order, third order and fifth order polynomial Kernel and answer the following questions.

10) From the following kernel functions, if you have to report one kernel function, which given **maximum classification error** on the training data. Which one will you report?

1 p

- second order polynomial
- third order polynomial
- fifth order polynomial
- all of the above options give same accuracy

No, the answer is incorrect.

Score: 0

Accepted Answers:

second order polynomial

11) For the IRIS data set, which kernel function among the following gave you the least number of support vectors?

1 point

- second order polynomial
- third order polynomial
- fifth order polynomial
- all of the above options give same number of support vectors

No, the answer is incorrect.

Score: 0

Accepted Answers:

third order polynomial

**Task 3:** After loading the IRIS data set, you have to learn SVM (SMO in WEKA). Your task is to learn SVM with RBF Kernel (with Gamma = 0.1) and answer the following question.

12) Which classifier among the three would you prefer to use for unseen data?

1 point

- second order polynomial
- third order polynomial
- fifth order polynomial
- RBF kernel with Gamma = 0.1

No, the answer is incorrect.

Score: 0

Accepted Answers:

third order polynomial

Previous Page

End

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

A project of



In association with



Funded by



Powered by

