

Name:



Reg No:

## APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

### 07 THRISSUR CLUSTER

#### SECOND SEMESTER M.TECH. DEGREE EXAMINATION APRIL 2018

#### DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

#### COMPUTER SCIENCE AND ENGINEERING

#### 07-CS-6106 MACHINE LEARNING AND LANGUAGE PROCESSING

Time:3 hours

Max.marks: 60

Answer all six questions. Part 'a' of each question is compulsory.

Answer either part 'b' or part 'c' of each question

<b>Q.no.</b>	<b>Module 1</b>	<b>Marks</b>
--------------	-----------------	--------------

1a	Describe the working of Naïve Byes classifier.	4
----	--	---

**Answer b or c**

b	Consider the following training data which is used for classification using decision tree. This data will be analysed and classified in to two classes, One is buys_ computer other is not buys_computer. What is Information gain? Find <b>Information gain</b> for the attributes age.	5
---	--	---

Rid	Age	Class: Buys_ Computer
1	Youth	No
2	Youth	No
3	Middle-aged	Yes
4	Senior	Yes
5	Senior	Yes
6	Senior	No
7	Middle-aged	Yes
8	Youth	No
9	Youth	Yes
10	Senior	Yes

c	Discuss in detail about Support Vector Machines and its applications.	5
---	---	---

Q.no.	Module 2	Marks
2a	What is Logistic regression? How we can do classification of data using this method?	4

**Answer b or c**

- b Suppose you are a marketing analyst for Disney toys. You gather the following data. After determining, via a scatter-plot, that the data followed a linear pattern, the regression line was found. What is the relationship between sales and advertisement?

Advertisement(x)	Sales(y)
2	3
4	7
6	5
8	10

- c Consider a scenario of online shopping website which sells different kind of magazines. As a review manager, you have to collect all reviews and needs to label them. Describe the idea of multi-label classification methods with this example.

Q.no.	Module 3	Marks
3a	Illustrate the idea of Expectation-Maximization algorithm.	4

**Answer b or c**

- b With the help of an example explain about K-means clustering in detail.
- c What is a sequential classifier? Write down the basic ideas of HMM and its applications in detail with necessary examples.

Q.no.	Module 4	Marks
4a	What are Maximum Entropy Models and its significance?	4

**Answer b or c**

- b Consider we had three urns urn 1, urn 2 and urn 3, these are three containers, and they have distributions of red, green and blue balls within them. 100 balls are in total for each urn. Number of balls in each urn is shown below.

BALLS	URN1(U1)	URN2 (U2)	URN3 (U3)
RED(R)	30	10	60
GREEN(G)	50	40	10
BLUE(B)	20	50	30

Find Transition probability matrix and observation likelihoods. Also draw the Viterbi tree for the above data.

- c Write a short note on Hidden Markov Model (HMM) and three problems. How the basic three problems of HMM can be solved? **5**

**Q.no. Module 5 Marks**

- 5a** Explain in detail about POS tagging and its applications.. **5**

**Answer b or c**

- b** Find out the relative frequency ( bigram probabilities) of the following corpus **7**

	I	want	to	study	Chinese
I	5	827	0	9	0
want	2	0	608	1	6
to	2	0	4	686	2
study	0	0	2	0	16
Chinese	1	0	0	0	0
for	15	0	15	0	1
an	2	0	0	0	0
interview	1	0	1	0	0

Consider the following vocabulary size  $V = 1416$  and Count of each word in the corpus is

I	want	to	study	Chinese
2533	927	2417	746	158

- c What is Named Entity Recognition (NER)? How NER can be created using statistical sequence labelling approach and just mention about evaluation measures of NER. **7**

**Q.no. Module 6 Marks**

- 6a** Explain about Factoid Question answering. How it is important for a search engine? **5**

**Answer b or c**

- b** Discuss in detail about the strategies used for machine translation. Explain with an example. **7**
- c** Discuss in detail about the architecture of a speech recognizer and some applications of Automatic Speech Recognition. **7**