USN ☐☐☐☐☐☐☐☐☐☐

PES Institute of Technology, Bangalore

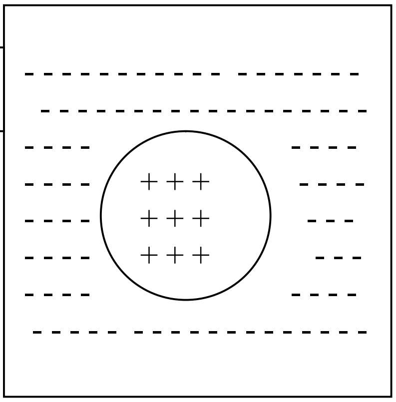
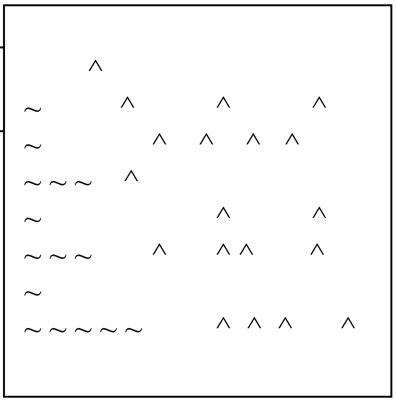
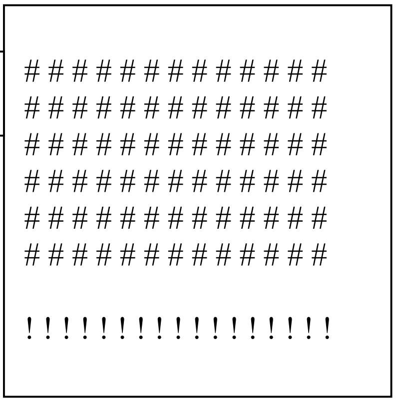
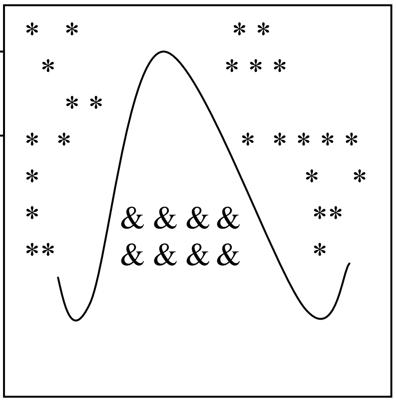
(Autonomous Institute under VTU, Belgaum)

CONTINUOUS INTERNAL EVALUATION (CIE) B. E. 6th SEMESTER

(Autonomous Even Semester) TEST – 1

12CS369: Applied Machine Learning

Time: 2 Hrs | Answer All Questions | Max Marks: 50

1.	a	Without any transformations to the input data, determine for which of the datasets depicted in the following graphs, the Perceptron Learning Algorithm will converge. (i)  (ii)  (iii)  (iv) 	2
1.	b	Given below is the Perceptron for OR gate in which x1 and x2 are features i.e., x1, x2={-1, +1} and w1 and w2 are weights. Hypothesis for the Perceptron representing OR gate is: **$h_w(x)=0.3+(0.5)x_1+(0.5)x_2$**	3

x_1	x_2	w_1x_1	w_2x_2	Σw_ix_i	w_0	$\Sigma w_ix_i + w_0$	$\text{sign}(\Sigma w_ix_i + w_0)$
-1	-1	-0.5	-0.5	-1	0.3	-0.7	0
-1	+1	-0.5	+0.5	0	0.3	0.3	1
+1	-1	+0.5	-0.5	0	0.3	0.3	1
+1	+1	+0.5	+0.5	+1	0.3	1.3	1

Design a Perceptron to implement $\overline{A} + \overline{B}$ and write the hypothesis for the same.

1. c Consider the set of training examples as below: 5

Instance	Class	A1	A2
1	+	T	T
2	+	T	T
3	-	T	F
4	+	F	F
5	-	F	T
6	-	F	T

(a) What is the entropy of training examples with respect to the target function classification?

(b) What is the information gain of the attribute A2 relative to the training examples?

2. a Consider the Sigmoidal function: 3

$g(z) = 1/(1+e^{-z})$

The derivative of sigmoidal function is used to compute and minimize errors in many Machine Learning hypotheses. Example: Neural Networks and Logistic Regression.

Prove that the derivative of the sigmoidal function is equal to **$g(z)\,(1-g(z))$**

2. b **tanh(x)** is a squashing function like Sigmoidal function which is used for classification in neural networks. A squashing function maps a large range of inputs to a narrow range. 3

$-1 <= \tanh(x) <= 1$

Given, **$\tanh(x)=(e^x-e^{-x})/(e^x+e^{-x})$**, prove that **$\tanh(x) = 2g(2x) – 1$**

2. c Consider a university which admits students based on the marks they have scored in their 12th standard examination. Students with a minimum mark of 60% and above are eligible to obtain admission in the university. The following table depicts admission details of the university obtained over the past three years. The first column shows the marks. The second 4

column is the number of students who have applied to the university and have the marks given in Column 1. The third column is the number of students who got admitted, out of the total number applied (Column 2), and having marks shown in Column 1.

Marks	Number of students who applied	No of students who got selected
60 - 70	100	1
70 - 80	100	8
80 - 90	100	18
90 - 92	100	30
93 - 94	100	45
95 -96	100	70
97 - 100	100	100

Model this as a Machine Learning problem in order to predict the probability of students getting selected. Please note that the sigmoidal function outputs a real – numbered value between 0 and 1, which can be interpreted as probability. Design a hypothesis which you think is a reflection of unknown target function. Also pictorially depict the hypothesis and prove that the curve is a valid probability distribution.

3. a Consider a MaxEnt classifier which classifies inputs into three classes: Person, Location and Drug. 8

The model, λ, is as follows: $\lambda_1 = 1.8$, $\lambda_2 = -0.6$, $\lambda_3 = 0.3$.

f_1, f_2, f_3 are feature functions or indicator functions which output a binary value.

The feature functions are given below:

f_1 (c,d) computes [c = Location $\wedge$ w_{-1} = “in” $\wedge$ isCapitalized(w)]

f_2 (c,d) computes [c = Location $\wedge$ hasAccentedLatinCharacter(w)]

f_3 (c,d) computes [c = Drug $\wedge$ ends(w,”c”)]

The equation to find the probability of occurrence of a class, given the document is:

$$\mathbf{P (c \mid d, \lambda) = (\exp \Sigma\ \lambda_i f_i (c,d)) / (\Sigma \exp \Sigma\ \lambda_i f_i (c,d))}$$

Given: Current Word = w (Goéric)
Previous Word = w_{-1} (by)

With reference to the above data, determine the following:

1. P(PERSON | by Goéric) =
2. P(LOCATION | by Goéric) =
3. P(DRUG | by Goéric) =

3. b Show that the equation specified in 3 (a) is a valid probability distribution. 2

USN

4. a

Suppose three students have built a binary classifier each. The classifiers are C1, C2 and C3. Out of the 1200 samples provided, 1000 are used for training and 200 are used for testing. The performance (number of samples correctly classified during training and testing) of the three classifiers is given below:

	C1	C2	C3
Training	975	900	700
Testing	100	175	130

With reference to the above data, answer the following questions:

(i) Which of these classifers have high bias?
(ii) Which of these classifers have high variance?
(iii) If you want to improve performance of the classifier which has high variance, what will you do?

4

4. b

Professor Ram is planning to publish a book consisting of 10 chapters. He has released a Beta copy of the book online. Each chapter of the book is a web page with hyperlinks to all other chapters. That is, the website hosting the book has the following structure:

(a) A home page that is accessed as root URL
(b) The home page describes the goal of the book and has links to all the 10 chapters. That is, there are 10 hyperlinks.
(c) Each chapter, which is in a separate web page, has 10 hyperlinks where each hyperlink navigates you to the respective chapter. Note that the current chapter is also listed as a hyperlink clicking which you stay with the same chapter.

Use the n-gram Language Model techniques to model the above scenario to design a solution to the following problems:

a) How would Prof Ram know the probability of a given page being visited?
b) Given that the user read chapters (i, j) in sequence, which chapter k he is likely to read next, where i, j, k are any chapters in the book?
c) If only four chapters can be read by a user due to a restriction that the author imposes, how many such 4 chapter sequences exist for ten chapters? For example, consider a user starting with reading a chapter i, followed by chapter j, followed by chapter k and chapter l; where i, j, k and l are any chapters.

6

5. a

The Naive Bayes Classifier makes assumptions. What are the assumptions made?

2

5.	b	Given a corpus of labeled tweets, design a sentiment analyzer using Naive Bayes Classifier which classifies the tweets into three states: Positive, Negative and Neutral. The tweet corpus contains tweet text and metadata. Metadata may include hashtags, user IDs, URLs, etc. What are the input features you will use and what are the model parameters and how will you estimate them?	6
5.	c	We have seen that linear models can be used to handle non – linearity by suitably transforming the inputs. Then, if this holds good, what is the purpose of true non – linear models?	2