National University of Computer and Emerging Sciences
School of Computing — Spring 2022 — Islamabad Campus

AI4001/CS4063 Fundamentals of NLP/NLP

Serial No:

Sessional II
Total Time: 1 Hour
Total Marks: 50

Usman.
Signature of Invigilator

Wednesday, May 11, 2022
Course Instructor
Mirza Omer Beg, Umair Arshad

Ali Kamal | 19I-1865 | N | @S
Student Name | Roll No | Section | Signature

DO NOT OPEN THE QUESTION BOOK OR START UNTIL INSTRUCTED.

Instructions:

1. Verify at the start of the exam that you have a total of four (4) questions printed on six (6) pages including this title page.
2. Attempt all questions on the question-book and in the given order.
3. This exam is **open book, open notes**. Mobiles, Internet and note-sharing is not allowed. Please see that the area in your threshold is free of any material classified as *useful in the paper*, i.e. mobile/internet or else there may be a charge of cheating.
4. Read the questions carefully for clarity of context and understanding of meaning and make assumptions wherever required, for neither the invigilator will address your queries, nor the teacher/examiner will come to the examination hall for any assistance.
5. Fit in all your answers in the provided space. You may use extra space on the last page if required. If you do so, clearly mark question/part number on that page to avoid confusion.
6. Use only your own stationery and calculator. If you do not have your own calculator, do manual calculations.
7. Use only permanent ink-pens. Only the questions attempted with permanent ink-pens will be considered. Any part of paper done in lead pencil cannot be claimed for checking/rechecking.

	Q1	Q2	Q3	Q4	Total
Marks Obtained	6	8	2	06	22
Total Marks	10	12	12	16	50

Q1. Short Questions (10 Marks) [10]

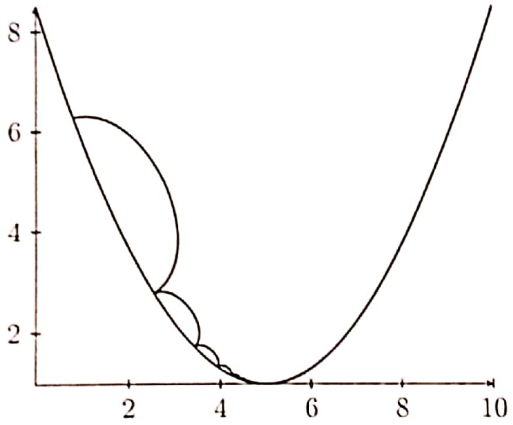
1. What does a logistic regressor compute?

   (a) An activation function followed by a linear function ($z = Wx + b$)
   (b) A linear function ($z = Wx + b$) followed by an activation function
   (c) A function g that scales the input x linearly ($Wx + b$)
   (d) The mean of all features before applying the output to an activation function
   (e) None of the above

2. Which of the following is the *Logistic Loss*?

   (a) $L(y, \hat{y}) = (y - \hat{y})^2$, (b) $L(y, \hat{y}) = |y - \hat{y}|$, (c) $L(y, \hat{y}) = -y \log(\hat{y}) - (1 - y) \log(1 - \hat{y})$
   (d) $L(y, \hat{y}) = -\frac{1}{m} \sum y \log(\hat{y}) - (1 - y) \log(1 - \hat{y})$ (e) None of the above

   Consider the parameter update equation $W', b' = W, b - \alpha \frac{d}{dW,b} L(y, \hat{y})$ and the gradient descent graph below.



3. There is a negative sign $(-)$ beside the learning rate (α) because:

   (a) Learning rate always decreases (b) Loss function always decreases
   (c) Parameters are always decreasing (d) All of the above (e) None of the above

4. What do the different sized jumps in the graph above represent?

   (a) Variable derivative of loss (b) Variable loss function (c) Variable learning rate
   (d) All of the above (e) None of the above

5. What is being represented by the x-axis of the graph above?

   (a) The parameters (b) The loss function (c) The learning rate
   (d) The derivative of the loss function (e) None of the above

National University of Computer and Emerging Sciences

School of Computing — Islamabad Campus

Q2. Text Classification (12 Marks) [8+4]

You are developing an emotion detection classifier that classifies sentence *affect* as *angry*(-), *calm*(=) or *cool*(+). Consider the following training corpus for the Multinomial Naive Bayes classifier with the given labels.

Training	
Sentence	Label
annoyed ~~by his~~ rage	-
~~very very~~ annoyed ~~at it~~	-
~~what does she do in her~~ rage	-
~~he is~~ cool	+
cool ~~and cool is~~ refreshing	+
calm ~~and calm it is~~	=

1. Considering that your classifier disregards stopwords={*what, very, and, he, his, her, she, by, is, at, it, in, do, does*}, compute priors and likelihood probabilities for the given classes.

Priors
(-): 3/6 , (+): 2/6 , (=): 1/6

|V| = 15

p(calm|-) = 1/9

with laplace smoothing ← Likelihood probabilities

P(annoyed|-) = 2+1/45 = 3/89 , P(rage|-) = 3/89 , P(cool|-) = 1/9 , P(refreshing|-) = 1/89

P(annoyed|+) = 0+1/315 = 1/8 , P(rage|+) = 1/8 , P(cool|+) = 3/8 , P(refreshing|+) = 2/8

P(calm|+) = 1/8

P(annoyed|=) = 0+1/1+5 = 1/6 , P(rage|=) = 1/6 , P(cool|=) = 1/6 , P(refreshing|=) = 1/6

P(calm|=) = 2/6

2. Classify the following test sentence. Show your work.

Test	Label
~~in her~~ rage ~~she is~~ calm ~~and~~ cool	?

→rage →calm →cool

(-) = 3/9 × 1/9 × 1/9 × (3/6) = 1/486 ← prior

(+) = 1/8 × 1/8 × 3/8 × (2/6) = 1/512

(=) = 1/6 × 2/6 × 1/6 × (1/6) = 1/648 ← prior

argmax of probabilities = label => (-)

Q3. Sentiment Analysis (12 Marks) [8+4]

The Multinomial Naïve Bayes pseudocode given below uses α-weighted Laplace smoothing to train a classification model. The algorithm uses words as features for classification.

Algorithm 1 TRAINMULTINOMIALNB($\mathcal{C}, \mathcal{T}$)

```
1: procedure NAIVEBAYESTRAINING(C, T)
2:     V ← EXTRACTVOCABULARY(C)
3:     N ← COUNTTEXTS(T)
4:
5:
6:     for each c ∈ C do
7:         N_c ← COUNTTEXTSINCLASS(T, c)
8:         N_w ← COUNTWORDSINALLTEXTSOFCLASS(T, c)
9:         prior[c] ← N_c / N
10:        doc_c ← CONCATENATETEXTSINCLASS(T, c)
11:        for each w_i ∈ V do
12:            N_i ← COUNTTOKENSOFWORDS(T)
13:        for each w_i ∈ V do
14:            condprob[w_i][c] ← (N_i + α) / (N_w + α|V|)
15:    return V, prior, condprob                    ▷ THE NB MODEL
```

if w_i in V do condprob[w_i][c] ← 1
else condprob[w_i][c] ← 0

(a) Modify the above code to convert it to the Boolean version of the Multinomial Naïve Bayes. A logical mistake has also been intentionally added to the above code. Circle the error and state the correction.

(b) The following shows the testing for the Multinomial Naïve Bayes classifier from the lectures. The probability scores are computed using $P(c).\prod_i P(w_i|c)$ may result in an underflow because of small probability values being computed. Suggest a solution and modify the algorithm given below to implement your solution.

use log to counter underflow

Algorithm 2 APPLYMULTINOMIALNB($\mathcal{C}, \mathcal{V}, prior, condprob, t$)

```
1: procedure NAIVEBAYESTEST(C, V, prior, condprob, t)
2:     W ← EXTRACTWORDSFROMTEXT(V, t)
3:     for each c ∈ C do
4:         score[c] ← prior[c]
5:         for each w_i ∈ W do
6:             score[c] *= condprob[w_i][c]
7:     return argmax_{c∈C} score[c]                 ▷ THE PREDICTED CLASS
```

score[c] += log(condprob[w_i][c])

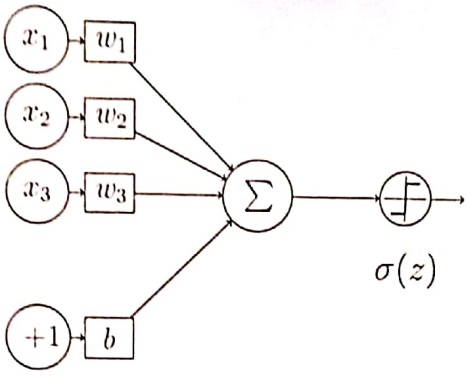
Q4. Logistic Regression

(16 Marks) [16]

Consider the two sentences given for for training the following Logistic Regressor.

s_1 = *Food was good, nice seating, soothing lighting.*

s_2 = *Bad taste, toxic environment. And to top it all off, unhygenic conditions!*



inputs weights

Given that s_1 is labeled positive and and s_2 is labeled as negative, and features x_1 (*count +ive lexicon*), x_2 (*count -ive lexicon*) and x_3 (*number of words*) and the initial weight and bias vector:

$$W, b = \begin{bmatrix} 1.5 \\ -0.25 \\ -1.0 \\ 0.3 \end{bmatrix} \quad (1)$$

Determine the weights and bias vector after training and backpropagating the logistic regressor on the above two sentences.

$\alpha = 0.1$

s_1 $y = 1$

Inference

$$\sigma\left(\begin{bmatrix} 1.5 \\ -0.25 \\ -1.0 \end{bmatrix}\begin{bmatrix} 3 \\ 0 \\ 7 \end{bmatrix} + 0.3\right) = \sigma(-2.8) = 0.099$$

label = 0, as $0.099 < 0.5$

updating weight

$$\begin{bmatrix} (\sigma(0)-1)(1.5) \\ (\sigma(0)-1)(-0.25) \\ (\sigma(0)-1)(-1) \\ (\sigma(0)-1) \end{bmatrix} \Rightarrow \begin{bmatrix} 1.5 \\ -0.25 \\ -1 \\ 0.3 \end{bmatrix} - 0.1 \begin{bmatrix} -0.75 \\ 0.125 \\ 0.5 \\ -0.5 \end{bmatrix} \times$$

$$= \begin{bmatrix} 1.575 \\ -0.26 \\ -1.05 \\ 0.35 \end{bmatrix} \Rightarrow$$ new weight and bias

Do the same for s_2 next page

0010

This page has been intentionally left empty.

Q2

$$x = \begin{bmatrix} 0 \\ 3 \\ 12 \end{bmatrix}$$