

National University of Computer and Emerging Sciences, Lahore Campus



Course: Data Analysis and Visualization
Program: BS (Data Science)
Duration: 180 Minutes
Paper Date: 17-Dec-2022
Section: A
Exam: Final Term

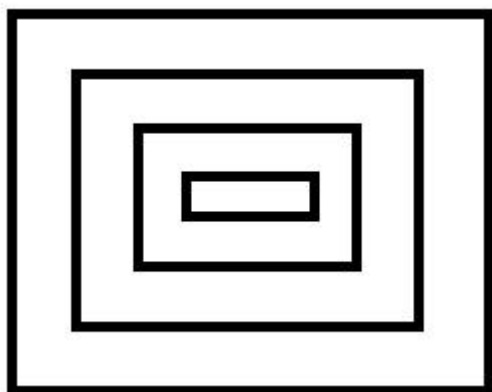
Course Code: DS-3001
Semester: Fall 2022
Total Marks: 85
Weight: 40%
Page(s): 12
Roll No.

Instruction/Notes: Attempt all questions. Answer in the space provided. You can ask for rough sheets but will not be attached with this exam. Answers written on rough sheet will not be marked. Do not use pencil or red ink to answer the questions. In case of confusion or ambiguity make a reasonable assumption.

Section 1 : (Short Conceptual Question Answer)

Q1 : What kind of filter is needed if I want to preserve vertical lines only in the given image? [3 Marks]

Input Image:



Filter Type : Derivative Filter

Filter Value:

$[-1 \ 0 \ 1]$

$[-1 \ 1]$

Any Horizontal derivative filter

Q2 : Which technique will you use if you want to select useful features only from the given dataset for better classification and to save time during training time? We have a subset of features given below. [2 Marks]

Dataset:

F1	F2	F3	F4	Output
2	1	4	2	Iris-setosa
4	3	8	6	Iris-setosa
6	7	12	14	Iris-versicolor
8	9	16	18	Iris-versicolor
10	11	20	22	Iris-versicolor

Technique:

Correlation Coefficient

Why:

Q3 : If we have a greater number of features, then which classifier will you use for better accuracy and why? for the following dataset? [2 Marks]

F1	F2	F3	F4	F5	F6	F30	Label
x	x	x	x	x	x	x	Y

Note: You have few 100 training examples

Classifier : SVM Classifier

Justification:

Q4 : For the given filters on input, what will be the output in the image given below? Origin is the middle pixel with value 5. Assume we have zero padding for corner cases.

A. In case of Cross Correlation what will be output? [3+3 Marks]

Stride = 1

Input:

0	0	0	0
0	0	0	0
0	0	1	0
0	0	0	0

Filter:

1	2	3
4	5	6
7	8	9

Output:

9	8	7	
6	5	4	
3	2	1	

B. In case of Convolution what will be output?

Input:

0	0	0	0
0	0	0	0
0	0	1	0
0	0	0	0

Filter:

1	2	3
4	5	6
7	8	9

Output:

1	2	3	
4	5	6	
7	8	9	

Q5: Which functionality in Plotly allows us to alter its internal detail. For example, color of its writings etc. [4 Marks]

Update Traces

Q6: What happens when we don't apply any kind of activation functions to the neurons in the neural network architecture? [5 Marks]

Note: Refer NN Architecture in Q1 of Long QA section on page#7

Discuss: (Generic Answers are also accepted in this question)

Non-Linearity will be vanished.

Q7: What restricts SVM to create its margins? Discuss any scenario with an example of where this could happen. [2 Marks]

Outliers

Q8: In Rectified Linear Unit (ReLU) Activation Function, a few neurons becomes dead during the training process.. [3 Marks]

A. What's reason behind it?

B. How to overcome this problem?

As it will zero / deactivate the value of negative outputs.
Leaky ReLU or Parameterized ReLU

Q9: You'll write an algorithm to classify whether images contain either a dog or a cat. This is easy for humans, dogs, and cats. Your computer will find it a bit more difficult. What will be the best choice of classifier for this kind of problem and all the related problems? [4 Marks]



Classifier you will use: CNN

Reason:

Convolutional NN is better for imagery datasets

Q10: If we have following Convolutional Neural Network (CNN) classifier then how many trainable parameters exist in this network. [8 Marks]

CNN Architecture Description:

1. We have a 28 x 28 RGB image having 8bit per pixel.
2. Then we have 5 convolutional filter of size 3 x 3. $\rightarrow 3 \times 3 \times 5$
3. Then we have pooling layer of window size 2 x 2.
4. Then we have 5 convolutional filter of size 5 x 5. $\rightarrow 5 \times 5 \times 5$
5. Then we have pooling layer of window size 2 x 2.
6. Then we have 5 neurons in hidden unit and a single neuron for output.

Output Dimensions

- 1 $28 \times 28 \times 3$
- 2 $26 \times 26 \times 5$
- 3 $13 \times 13 \times 5$
- 4 $9 \times 9 \times 5$
- 5 $5 \times 5 \times 5$
- 6 $4 \times 4 \times 5$
- 7 $2 \times 2 \times 5 \times 5 + 5 + 1$
- 8 $5 \times 5 \times 5 \times 5 + 5 + 1$

Note: Write your answer in terms of decimal number only.

Trainable Parameters: _____

$$45 + 125 + 400 + 5 + 1 = 576 \quad / \quad 45 + 125 + 625 + 5 + 1 = 801$$

$5 \cdot 1$ $5 \cdot 2$

Q11: Which feature extraction technique do you use in textual data where semantic sequence is not important for the given dataset and why? The Spam and No Spam mail dataset's subset is given below. [3 Marks]

Note: As discussed in class.

Input 1: Hello! Ali how are you.	Feature Extractor Name: <u>BoW or TF-IDF</u> Justification:
Input 2: You got 50 Lac Dollar in a lucky draw.	
Input 3: You number won a Iphone 14 pro max this weekend. Claim it via link.	
Input 4: A property owner died and your details lucky matched. Please contact.	
Input 5: Hello Sir Hope you are doing fine.	
Input 6: Your Credential are required to get you a brand new AUDI.	

Q12: Which feature extraction technique do you use in textual data where semantic sequences are important for the given dataset and why? A product review on tablets and laptops subset is given below. [3 Marks]

Note: As discussed in class.

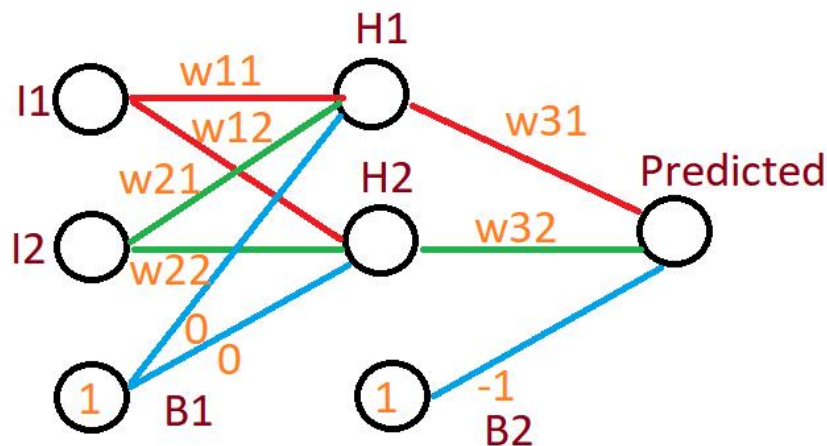
<p>Input 1: Samsung's 8 inch tablet are best.</p> <p>Input 2: Oppo introduces its best 8 inch OLED display.</p> <p>Input 3: Apple come with better features in 8 inch category, your next laptop.</p> <p>Input 4: Google Pixel is gonna impact market by its new 8 inch pixel tablet.</p> <p>Input 5: Nokia's 8 inch tablets are gonna save their market repute.</p> <p>Input 6: Tablets are good but not as good as laptops are.</p>	<p>Feature Extractor Name: <u>N-Grams</u></p> <p>Justification:</p>
---	---

Section 2 : (Long Solvable Question Answer)

Q1: XOR Gate is the prime example of non-linearity as its data is not linearly separable. So here comes the need for a neural network packed with multiple neurons and also multiple layers are stacked on one another to learn a higher degree of patterns. In this task you are required to fill in the table given below where we have 2 dimensional XOR input along with a bias value and the actual label is also provided for comparison. [20 Marks]

I1	I2	W11	W12	W21	W22	B1	H1	H2	W31	W32	B2	Predicted	Actual
0	0	1	1	1	1	0	0	0	1	1	-1	0.27	0
0	1	1	1	1	1	0	1	1	1	1	-1	0.73	1
1	0	1	1	0.81	1	0	1	1	0.99	0.99	-0.92	0.72	1
1	1	1	1	0.81	1	0	1.81	2	0.99	0.99	-0.92	0.93	0

Note: Here you are supposed to use mini batch gradient decent with mini batch size = 2. The activation unit is ReLU in the case of hidden units, while sigmoid. The learning rate is 0.7. Compute forward pass and backward pass for the scenario given and fill out the values in the given table. Actual - Predicted is the last layer Error Signal and for the hidden unit you have to take a derivative of ReLU, which is if the hidden unit value is 1, then consider the error or otherwise neglect it. Moreover, the Architecture Diagram is given below for more in-depth understanding.



Required Work for NN

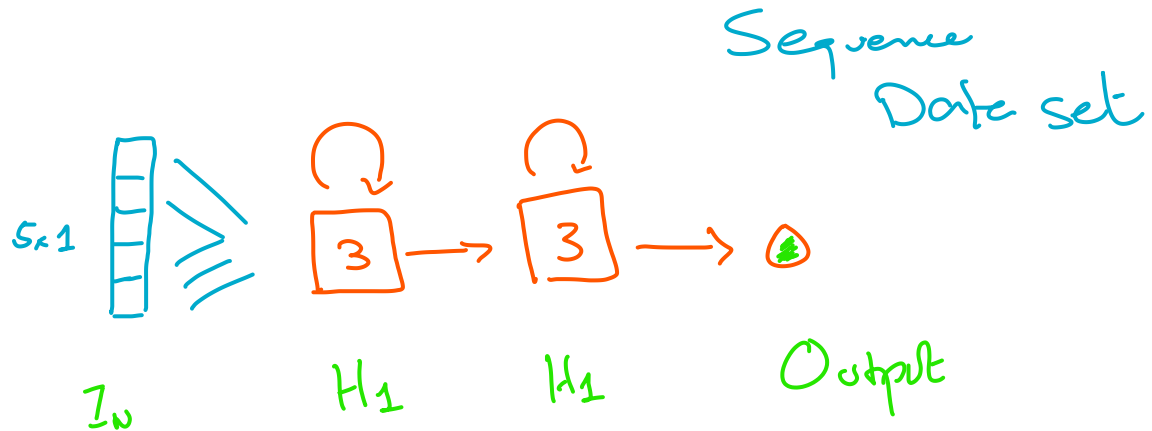
Q2: You are required to draw the architectural diagram for the RNN classifier. [10 Marks]

Architecture Details:

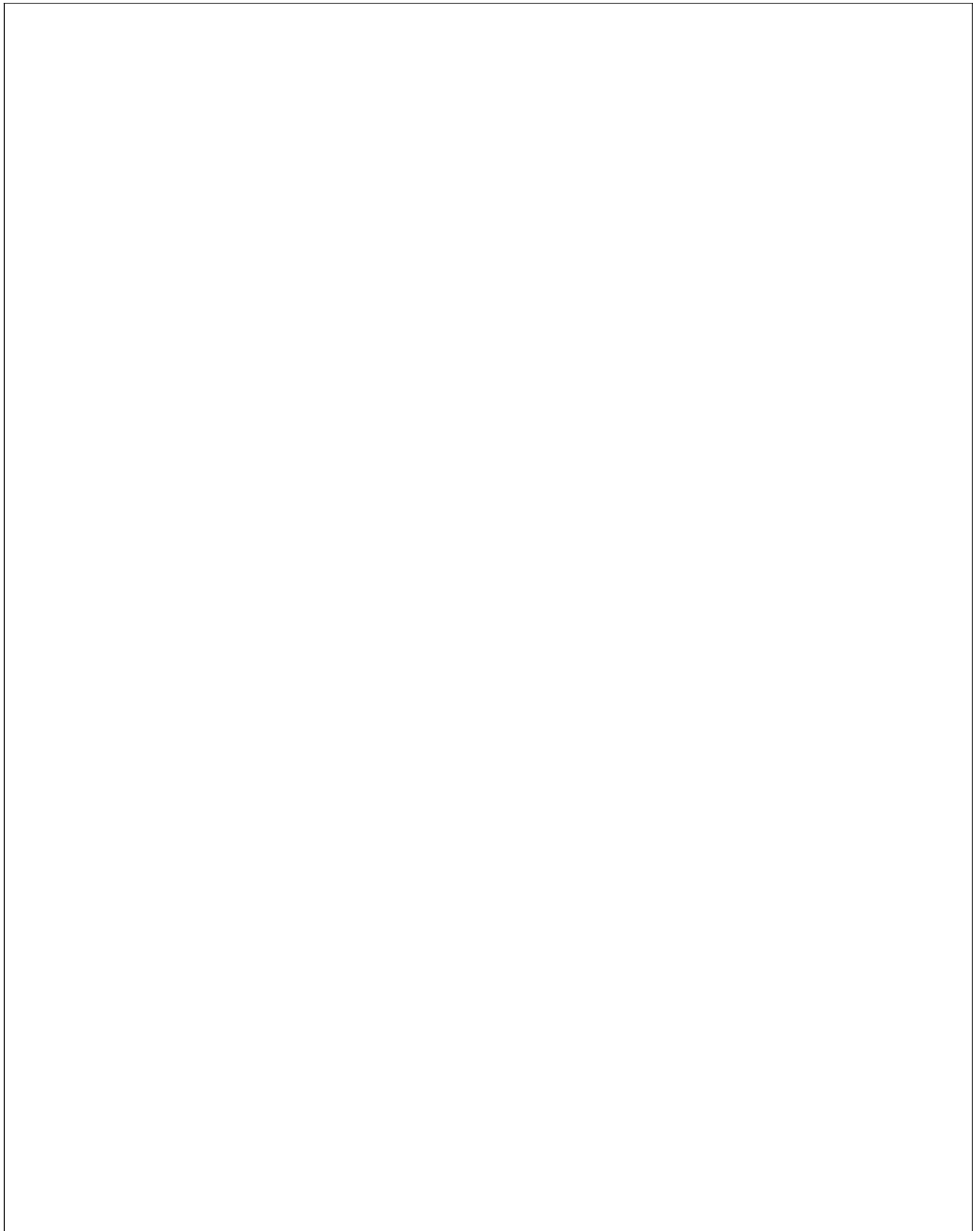
- Input unit contains 5 dimensional feature vector.
- Hidden layer 1 contains 3 neurons.
- Hidden Layer 2 contains 3 neurons.
- Output Layer Contains only 1 neuron for output.

As every input line feeds into this network contains exactly 2 words, you require to draw its architecture for time t_0 and t_1 . t_0 for 1st word input while t_1 for 2nd word input before producing a output. Words could be anything and labels could be anything. Your task is to produce a diagrammatic view of the whole architecture only.

Abstract Level Diagram for RNN.



Detailed Diagram with respect to time t_0 and t_1 .



Q3: For the scenario given below, what type of interactive graphing technique you will use and what your plotting questions will be. How you make it interactive list down your with selected attributes choices.
[10 Marks]

Indian Prisoner Statistics:

This dataset contains the complete details about the prison and various characteristics of inmates. This will help you to understand better about the prison system in India. The following Attributes are taken under consideration.

State Name	Year	Gender	Caste	Under Trail	Convicts	Detenues
------------	------	--------	-------	-------------	----------	----------

There are 7 unique states named a,b,c,d,e,f and the dataset is available from the year 2003 to 2011. Gender is classified as male and female. The cast has 4 unique values, I.e., w,x,y,z. and numbers are placed in the last 3 columns to keep the count of prisoners. A subset taken from the original dataset is in the figure below for more intuition about the scenario.

state_name	year	gender	caste	convicts	under_trial	detenues
Andhra Pradesh	2001	Male	SC	1236	1836	0
Andhra Pradesh	2001	Male	ST	418	1205	0
Arunachal Pradesh	2012	Male	ST	3	28	0
Arunachal Pradesh	2012	Male	OBC	0	5	0
Arunachal Pradesh	2012	Male	Others	1	26	0
Assam	2001	Male	SC	278	548	1
Bihar	2012	Male	ST	187	1449	0
Bihar	2012	Male	OBC	2036	9904	14
Bihar	2012	Male	Others	1223	7023	24
Chhattisgarh	2011	Male	OBC	1856	2428	0
Chhattisgarh	2011	Male	Others	567	1039	0
Goa	2001	Male	SC	6	23	0
Goa	2001	Male	ST	3	9	0
Gujarat	2004	Male	Others	781	1963	197
Gujarat	2005	Male	SC	916	1523	89
Gujarat	2005	Male	ST	977	1524	46
Gujarat	2005	Male	OBC	1456	1656	80

Solution: [Descriptive Answer required only]

Bar Chart → • Update Values