Register Number:……..

KUMARAGURU
college of technology
character is life

B.E/B.TECH DEGREE EXAMINATIONS: APRIL / MAY 2023

(Regulation 2018)

Sixth Semester

COMMON TO ALL BRANCHES

U18MAR0204: Deep Learning

COURSE OUTCOMES

CO1: Understanding the working of Neural Networks
CO2: Applying the deep learning concepts for real-time case studies

Time: Three Hours **Maximum Marks: 100**

Answer all the Questions:-

PART A (10 x 2 = 20 Marks)

(Answer not more than 40 words)

1.	What is the gradient of the following function at x = y = 0? $f(x,y) = (x-2)^3 \sin(y)$	CO1	[K3]
2.	Describe one way a neural network could be initialized poorly and what effect that could have.	CO1	[K2]
3.	After training a neural network, you observe a large gap between the training accuracy (100%) and the test accuracy (42%). Which method is commonly used to reduce this gap?	CO1	[K3]
4.	Is the following convolution kernel separable? If so, separate it. If not, prove that it is not. $H = \begin{array}{\|c\|c\|}\hline 2 & 3 \\ \hline 1 & 1 \\ \hline \end{array}$	CO1	[K2]
5.	Describe two difficulties that computer vision algorithms face in dealing with images. i.e. two characteristics of image formation that make it difficult to recover the image content.	CO1	[K2]
6.	Give two benefits of using convolutional layers instead of fully connected ones for visual tasks.	CO2	[K3]
7.	Consider, in a convolution neural network we supply 32x32x3 input volume along with ten 5x5 filters with stride=1 and pad=2, what is the number of parameters in this layer?	CO2	[K2]
8.	Compare RNN with LSTM.	CO2	[K2]
9.	Sketch a simple recurrent network, with input x, output y, and recurrent state h.	CO2	[K2]
10.	You are designing a deep learning system to detect driver fatigue in cars. It is crucial that that your model detects fatigue, to prevent any accidents. Which of the following is the most appropriate evaluation metric: Accuracy, Precision, Recall, Loss Value? Justify.	CO2	[K3]

Answer any FIVE Questions:-
PART B (5 x 16 = 80 Marks)
(Answer not more than 400 words)

11. Assume we have a set of data from patients who have visited Kumaraguru hospital during the year 2021. A set of features (e.g., temperature, height) have been also extracted for each patient. Our goal is to decide whether a new visiting patient has any of diabetes, heart disease, or Alzheimer (a patient can have one or more of these diseases).

a) We have decided to use a neural network to solve this problem. We have two choices: either to train a separate neural network for each of the diseases or to train a single neural network with one output neuron for each disease, but with a shared hidden layer. Which method do you prefer? Justify your answer. 8 CO1 [K4]

b) Some patient features are expensive to collect (e.g., brain scans) whereas others are not (e.g., temperature). Therefore, we have decided to first ask our classification algorithm to predict whether a patient has a disease, and if the classifier is 80% confident that the patient has a disease, then we will do additional examinations to collect additional patient features. In this case, which classification methods do you recommend? Justify your answer. 8 CO1 [K4]

12. a) Consider a CNN classifier. For each layer, calculate the number of weights, number of biases and the size of the associated feature maps. 8 CO1 [K4]

The notation follows the convention:

• CONV-K-N denotes a convolutional layer with N filters, each them of size K×K, Padding and stride parameters are always 0 and 1 respectively.

• POOL-K indicates a K × K pooling layer with stride K and padding 0.

• FC-N stands for a fully-connected layer with N neurons.

Layer	Activation map dimensions	Number of weights	Number of biases
INPUT	128 × 128 × 3	0	0
CONV-9-32			
POOL-2			
CONV-5-64			
POOL-2			
CONV-5-64			
POOL-2			
FC-3			

	b)	The loss functions of deep networks are non-convex and there was concern in the early days of the field that gradient methods may get trapped in poor (i.e. high loss) local optima. Describe with a sketch how local optima were observed to behave on a simple model, then explain why SGD performs well in practice on such models.	8	CO1	[K4]
13.	a)	Give a full description of the Convolutional layer with suitable formulae and diagram.	8	CO1	[K2]
	b)	The convolution of a 2D image f(x, y) and a kernel h(x, y) is defined as	8	CO1	[K3]

$$g(x,y) = (f * h)(x,y) = \sum_{\alpha=-\infty}^{\infty} \sum_{\beta=-\infty}^{\infty} f(\alpha,\beta)h(x-\alpha, y-\beta)$$

Perform the convolution below, i.e. calculate the image C. All values outside the image array A are equal to zero. In the arrays A and B, the respective number written in bold face is at position (x, y) = (0, 0). Note that C is only a part of the convolution result.

0	1	0	2		0	1	0		**?**	?	?	?
1	0	1	0	*	1	**1**	2	=	?	?	?	?
0	2	0	0		0	2	0		?	?	?	?
image A				*	kernel B			=	image C			

14.	a)	Explain the different hyper-parameters that define a convolutional layer.	8	CO2	[K2]
	b)	Compare LeNets and AlexNets. What are the main architectural features of each, and how did they lead to improvements over previous design? Use diagrams of the networks as appropriate	8	CO2	[K2]
15.	a)	Consider a recurrent neural network (RNN) takes in an input vector x_t and a state vector h_{t-1} and returns a new state vector h_t and an output vector y_t:	8	CO2	[K3]

$$h_t = f(w_1 x_t + w_2 h_{t-1} + b_2)$$
$$y_t = g(w_3 h_t + b_3),$$

The following diagram depicts a single RNN unit, where x_t, h_{t-1}, h_t and y_t are all scalars as a state machine:

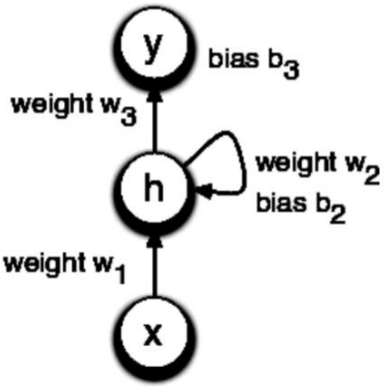
$w_1 =$ _____

$w_2 =$ _____

$b_2 =$ _____

$w_3 =$ _____

$b_3 =$ _____



Suppose that f is a binary threshold unit and g is a linear unit:

$$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases}$$

$$g(x) = x.$$

Fill in weights (w_1, w_2, w_3), biases (b_1, b_2) so that the RNN initially outputs 0, but as soon as it receives an input of 1, it switches to outputting 1 for all subsequent time steps. For instance, the input 0001010 produces the output 0001111. The hidden unit has an initial value of 0.

b) Suppose you have a dataset with N samples. You would like to train a deep network and tune learning rate, minibatch size and regularization to give the best model. Describe the cross-validation design you would use to tune the hyper-parameters, and produce the most accurate possible (unbiased) estimate of the model's test loss. i.e. describe how to partition the data, and what to do with each partition. 8 CO2 [K3]

16. a) How LSTM networks solve the problem of vanishing gradients? Explain. 8 CO2 [K2]

b) Briefly contrast the backward updates for a ReLU layer using (i) normal backpropagation, (ii) guided backpropagation and (iii) deconvolution 8 CO2 [K2]

****