CST 395	NEURAL NETWORKS AND DEEP LEARNING	Category	L	T	P	Credit	Year of Introduction
		VAC	3	1	0	4	2019

Preamble:

Neural networks is a biologically inspired programming paradigm which enables a computer to learn from observational data and deep learning is a powerful set of techniques for training neural networks. This course introduces the key concepts in neural networks, its architecture and learning paradigms, optimization techniques, basic concepts in deep learning, Convolutional Neural Networks and Recurrent Neural Networks. The students will be able to provide best solutions to real world problems in domains such as computer vision and natural language processing.

Prerequisite: A Sound knowledge in Computational fundamentals of machine learning

Course Outcomes: After the completion of the course the student will be able to

CO1	Demonstrate the basic concepts of machine learning models and performance measures. **(Cognitive Knowledge Level : Understand)**
CO2	Illustrate the basic concepts of neural networks and its practical issues**(Cognitive Knowledge Level : Apply)**
CO3	Outline the standard regularization and optimization techniques for deep neural networks (**Cognitive Knowledge Level : Understand)**
CO4	Build CNN and RNN models for different use cases. **(Cognitive Knowledge Level : Apply)**
CO5	Explain the concepts of modern RNNs like LSTM, GRU **(Cognitive Knowledge Level : Understand)**

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	✓	✓	✓	✓								✓
CO2	✓	✓	✓	✓								✓
CO3	✓	✓	✓	✓								✓
CO4	✓	✓	✓	✓	✓	✓						✓
CO5	✓	✓	✓	✓								✓

Abstract POs defined by National Board of Accreditation			
PO#	**Broad PO**	**PO#**	**Broad PO**
PO1	Engineering Knowledge	**PO7**	Environment and Sustainability
PO2	Problem Analysis	**PO8**	Ethics
PO3	Design/Development of solutions	**PO9**	Individual and team work
PO4	Conduct investigations of complex problems	**PO10**	Communication
PO5	Modern tool usage	**PO11**	Project Management and
PO6	The Engineer and Society	**PO12**	Life long learning

Assessment Pattern

Bloom's Category	Continuous Assessment Tests		End Semester Examination Marks
	Test1 (%)	Test2 (%)	
Remember	30	30	30
Understand	40	40	40
Apply	30	30	30
Analyse			
Evaluate			
Create			

Mark Distribution

Total Marks	CIE Marks	ESE Marks	ESE Duration
150	50	100	3 hours

Continuous Internal Evaluation Pattern

Attendance	**10 marks**
Continuous Assessment Tests	**25 marks**
Continuous Assessment Assignment	**15 marks**

Internal Examination Pattern:

Each of the two internal examinations has to be conducted out of 50 marks. First Internal Examination shall be preferably conducted after completing the first half of the syllabus and the Second Internal Examination shall be preferably conducted after completing the remaining part of the syllabus. There will be two parts: Part A and Part B. Part A contains 5 questions (preferably, 2 questions each from the completed modules and 1 question from the partly covered module), having 3 marks for each question adding up to 15 marks for part A. Students should answer all questions from Part A. Part B contains 7 questions (preferably, 3 questions each from the completed modules and 1 question from the partly covered module), each with 7 marks. Out of the 7 questions in Part B, a student should answer any 5.

End Semester Examination Pattern:

There will be two parts; Part A and Part B. Part A contains 10 questions with 2 questions from each module, having 3 marks for each question. Students should answer all questions. Part B contains 2 questions from each module of which a student should answer any one. Each question can have a maximum 2 subdivisions and carry 14 marks.

Syllabus

Module - 1 (Basics of Machine Learning)

Machine Learning basics - Learning algorithms - Supervised, Unsupervised, Reinforcement, Overfitting, Underfitting, Hyper parameters and Validation sets, Estimators -Bias and Variance. Challenges in machine learning. Simple Linear Regression, Logistic Regression, Performance measures - Confusion matrix, Accuracy, Precision, Recall, Sensitivity, Specificity, Receiver Operating Characteristic curve(ROC), Area Under Curve(AUC).

Module -2 (Neural Networks)

Introduction to neural networks -Single layer perceptrons, Multi Layer Perceptrons (MLPs), Representation Power of MLPs, Activation functions - Sigmoid, Tanh, ReLU, Softmax. Risk minimization, Loss function, Training MLPs with backpropagation, Practical issues in neural network training - The Problem of Overfitting, Vanishing and exploding gradient problems, Difficulties in convergence, Local and spurious Optima, Computational Challenges. Applications of neural networks.

Module 3 (Deep learning)

Introduction to deep learning, Deep feed forward network, Training deep models, Optimization techniques - Gradient Descent (GD), GD with momentum, Nesterov accelerated GD, Stochastic GD, AdaGrad, RMSProp, Adam. Regularization Techniques - L1 and L2 regularization, Early stopping, Dataset augmentation, Parameter sharing and tying, Injecting noise at input, Ensemble methods, Dropout, Parameter initialization.

Module -4 (Convolutional Neural Network)

Convolutional Neural Networks – Convolution operation, Motivation, Pooling, Convolution and Pooling as an infinitely strong prior, Variants of convolution functions, Structured outputs, Data types, Efficient convolution algorithms. Practical use cases for CNNs, Case study - Building CNN model AlexNet with handwritten digit dataset MNIST.

Module- 5 (Recurrent Neural Network)

Recurrent neural networks – Computational graphs, RNN design, encoder – decoder sequence to sequence architectures, deep recurrent networks, recursive neural networks, modern RNNs LSTM and GRU, Practical use cases for RNNs. Case study - Natural Language Processing.

Text Book

1. Goodfellow, I., Bengio,Y., and Courville, A., Deep Learning, MIT Press, 2016.
2. Neural Networks and Deep Learning, Aggarwal, Charu C., c Springer International Publishing AG, part of Springer Nature 2018
3. Fundamentals of Deep Learning: Designing Next-Generation Machine Intelligence Algorithms (1st. ed.). Nikhil Buduma and Nicholas Locascio. 2017. O'Reilly Media, Inc.

Reference Books

1. Satish Kumar, Neural Networks: A Classroom Approach, Tata McGraw-Hill Education, 2004.
2. Yegnanarayana, B., Artificial Neural Networks PHI Learning Pvt. Ltd, 2009.
3. Michael Nielsen, Neural Networks and Deep Learning, 2018

Course Level Assessment Questions

Course Outcome 1 (CO1):

1. Predict the price of a 1000 square feet house using the regression model generated from the following data.

No.	Square feet	Price(Lakhs)
1	500	5
2	900	10
3	1200	13
4	1500	18
5	2000	25
6	2500	32
7	2700	35

2. Consider a two-class classification problem of predicting whether a photograph contains a man or a woman. Suppose we have a test dataset of 10 records with expected outcomes and a set of predictions from our classification algorithm. Compute the confusion matrix, accuracy, precision, recall, sensitivity and specificity on the following data.

Sl.**No.**	Actual	Predicted
1	man	woman
2	man	man
3	woman	woman
4	man	man

5	man	woman
6	woman	woman
7	woman	man
8	man	man
9	man	woman
10	woman	woman

Course Outcome 2 (CO2):

1. Suppose you have a 3-dimensional input x = (x1, x2, x3) = (2, 2, 1) fully connected with weights (0.5, 0.3, 0.2) to one neuron which is in the hidden layer with sigmoid activation function. Calculate the output of the hidden layer neuron.
2. Consider the case of the XOR function in which the two points {(0, 0),(1, 1)} belong to one class, and the other two points {(1, 0),(0, 1)} belong to the other class. Design a multilayer perceptron for this binary classification problem.

Course Outcome 3 (CO3):

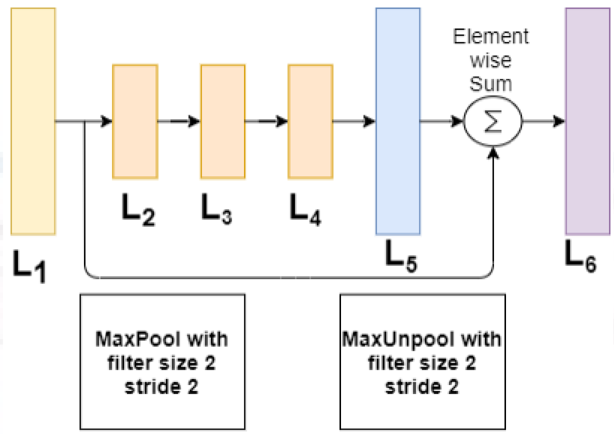
1. Derive a mathematical expression to show L2 regularization as weight decay.
2. Explain how L2 regularization improves the performance of deep feed forward neural networks.
3. Explain how L1 regularization method leads to weight sparsity.

Course Outcome 4 (CO4):

1. Draw and explain the architecture of convolutional neural networks.
2. You are given a classification problem to classify the handwritten digits. Suggest a learning and/or inference machine with its architecture, an objective function, and an optimization routine, along with how input and output will be prepared for the classifier.

3. In a Deep CNN architecture the feature map L_1 was processed by the following operations as shown in the figure. First down sampled using max pool operation of size 2 and stride 2, and three convolution operations and finally max unpool operation and followed by an element wise sum. The feature map L_1 and L_4 are given below. Compute the matrix L6.

$$L_1 = \begin{matrix} 10 & 20 & 15 & 22 \\ 20 & 16 & 28 & 30 \\ 30 & 12 & 20 & 16 \\ 20 & 20 & 40 & 12 \end{matrix} \qquad L_4 = \begin{matrix} 10 & 20 \\ 20 & 30 \end{matrix}$$



4. Illustrate the workings of the RNN with an example of a single sequence defined on a vocabulary of four words.

Course Outcome 5 (CO5):

1. Draw and explain the architecture of LSTM.
2. List the differences between LSTM and GRU

Model Question Paper

QP CODE: **PAGES:4**

Reg No:_______________

Name:_________________

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

FIFTH SEMESTER B.TECH DEGREE EXAMINATION(HONORS), MONTH & YEAR

Course Code: CST 395

Course Name: Neural Networks and Deep Learning

Max.Marks:100 **Duration:3 Hours**

PART A

Answer all Questions. Each question carries 3 Marks

1. List and compare the types of machine learning algorithms
2. Suppose 10000 patients get tested for flu; out of them, 9000 are actually healthy and 1000 are actually sick. For the sick people, a test was positive for 620 and negative for 380. For healthy people, the same test was positive for 180 and negative for 8820. Construct a confusion matrix for the data and compute the

accuracy, precision and recall for the data

3. Illustrate the limitation of a single layer perceptron with an example

4. Specify the advantages of ReLU over sigmoid activation function.

5. Derive weight updating rule in gradient descent when the error function is a) mean squared error b) cross entropy

6. List any three methods to prevent overfitting in neural networks

7. What happens if the stride of the convolutional layer increases? What can be the maximum stride? Justify your answer.

8. Consider an activation volume of size 13×13×64 and a filter of size 3×3×64. Discuss whether it is possible to perform convolutions with strides 2, 3 and 5. Justify your answer in each case.

9. How does a recursive neural network work?

10. List down three differences between LSTM and RNN

(10x3=30)

Part B

(Answer any one question from each module. Each question carries 14 Marks)

11. (a) Prove that the decision boundary of binary logistic regression is linear **(9)**

(b) Given the following data, construct the ROC curve of the data. Compute the AUC. **(5)**

Threshold	**TP**	TN	FP	FN
1	0	25	0	29
2	7	25	0	22
3	18	24	1	11
4	26	20	5	3
5	29	11	14	0

6	29	0	25	0
7	29	0	25	0

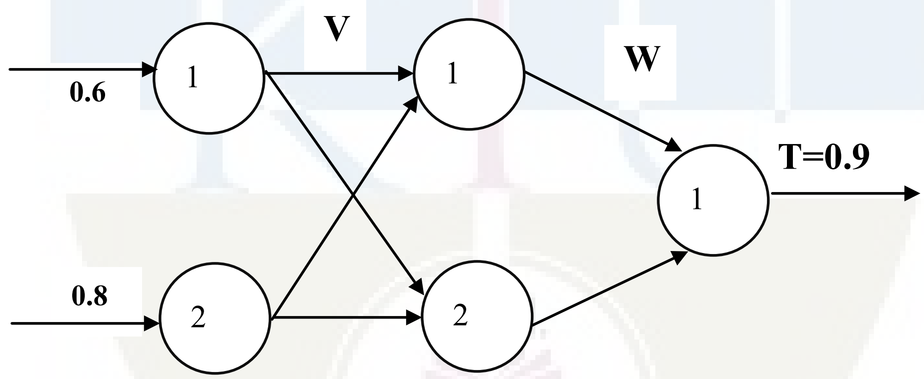
OR

12. (a) With an example classification problem, explain the following terms:
a) Hyper parameters b) Training set c) Validation sets d) Bias e) Variance **(8)**

(b) Determine the regression equation by finding the regression slope coefficient and the intercept value using the following data. **(6)**

x	55	60	65	70	80
y	52	54	56	58	62

13. (a) Update the parameters V_{11} in the given MLP using back propagation with learning rate as 0.5 and activation function as sigmoid. Initial weights are given as V_{11}= 0.2, V_{12}=0.1, V_{21}=0.1, V_{22}=0.3, V_{11}=0.2, W_{11}=0.5, W_{21}=0.2 **(10)**

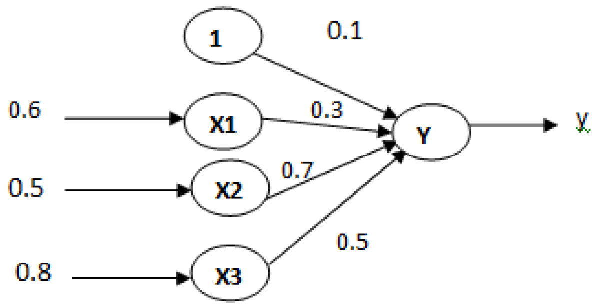


(b) Explain the importance of choosing the right step size in neural networks **(4)**

OR

14. (a) Explain in detail any four practical issues in neural network training **(8)**

(b) Calculate the output of the following neuron Y with the activation function as a) binary sigmoid b) tanh c)ReLU **(6)**



15. (a) Explain, what might happen in ADAGRAD, where momentum is expressed as $\Delta \square_{\square} = -\square\square_{\square}/\sqrt{(\sum_{\square=1}^{\square} \square_{\square}^{2})}$ where the denominator computes the L2 norm of all previous gradients on a per-dimension basis and □ is a global learning rate shared by all dimensions. **(6)**

(b) Differentiate gradient descent with and without momentum. Give equations for weight updation in GD with and without momentum. Illustrate plateaus, saddle points and slowly varying gradients. **(8)**

OR

16. (a) Suppose a supervised learning problem is given to model a deep feed forward neural network. Suggest solutions for the following a) small sized dataset for training b) dataset with both labelled and unlabeled data c) large data set but data from different distribution **(9)**

(b) Describe the effect in bias and variance when a neural network is modified with more number of hidden units followed with dropout regularization. **(5)**

17. (a) Draw and explain the architecture of Convolutional Neural Networks **(8)**

(b) Suppose that a CNN was trained to classify images into different categories. It performed well on a validation set that was taken from the same source as the training set but not on a testing set. What could be the problem with the training of such a CNN? How will you ascertain the problem? How can those problems be solved? **(6)**

OR

18. (a) Explain the following convolution functions a)tensors b) kernel flipping c) down sampling d) strides e) zero padding. **(10)**

	(b)	What is the motivation behind convolution neural networks?	**(4)**
19.	(a)	Describe how an LSTM takes care of the vanishing gradient problem. Use some hypothetical numbers for input and output signals to explain the concept	**(8)**
	(b)	Explain the architecture of Recurrent Neural Networks	**(6)**

OR

20.	(a)	Explain LSTM based solution for anyone of the problems in the Natural Language Processing domain.	**(8)**
	(b)	Discuss the architecture of GRU	**(6)**

Teaching Plan

Module 1 : [Text book 1: Chapter 5, Textbook 2: Chapter 2](9 hours)		
1.1	Introduction, Learning algorithms - Supervised, Unsupervised, Reinforcement	1 hour
1.2	Overfitting, Underfitting, Hyperparameters	1 hour
1.3	Validation sets, Estimators -Bias and Variance. Challenges in machine learning.	1 hour
1.4	Simple Linear Regression	1 hour
1.5	Illustration of Linear Regression	1 hour
1.6	Logistic Regression	1 hour
1.7	Illustration of Logistic Regression	1 hour
1.8	Performance measures - Confusion matrix, Accuracy, Precision, Recall, Sensitivity, Specificity, ROC, AUC.	1 hour
1.9	Illustrative Examples for performance measures	1 hour
Module 2 : Text book 2, Chapter 1 (8 hours)		
2.1	Introduction to neural networks -Single layer perceptrons	1 hour
2.2	Multi Layer Perceptrons (MLPs), Representation Power of MLPs	1 hour
2.3	Activation functions - Sigmoid, Tanh, ReLU, Softmax. Risk minimization, Loss function	1 hour

2.4	Training MLPs with backpropagation	1 hour
2.5	Illustration of back propagation algorithm	1 hour
2.6	Practical issues in neural network training - The Problem of Overfitting, Vanishing and exploding gradient problems	1 hour
2.7	Difficulties in convergence, Local and spurious Optima, Computational Challenges.	1 hour
2.8	Applications of neural networks	1 hour
Module 3 : Text book 1: Chapter 7, 8, Text book 2, Chapter 3, 4 (10 hours)		
3.1	Introduction to deep learning, Deep feed forward network	1 hour
3.2	Training deep models - Introduction, setup and initialization issues	1 hour
3.3	Solving vanishing and exploding gradient problems	1 hour
3.4	Concepts of optimization, Gradient Descent (GD), GD with momentum.	1 hour
3.5	Nesterov accelerated GD, Stochastic GD.	1 hour
3.6	AdaGrad, RMSProp, Adam.	1 hour
3.7	Concepts of Regularization, L1 and L2 regularization.	1 hour
3.8	Early stopping, Dataset augmentation	1 hour
3.9	Parameter sharing and tying, Injecting noise at input, Ensemble methods	1 hour
3.10	Dropout, Parameter initialization.	1 hour
Module 4 : Text book 1, Chapter 9, Text book 2: Chapter 8 (8 hours)		
4.1	Convolutional Neural Networks, architecture	1 hour
4.2	Convolution and Pooling operation with example	1 hour
4.3	Convolution and Pooling as an infinitely strong prior	1 hour
4.4	Variants of convolution functions, structured outputs, data types	1 hour
4.5	Efficient convolution algorithms.	1 hour
4.6	Practical use cases for CNNs	1 hour
4.7	Case study - Building CNN with MNIST and AlexNet.	1 hour
4.8	Case study - Building CNN with MNIST and AlexNet	1 hour
Module 5 : Text book 1 :Chapter 10, 11, Text book 2:Chapter 7 (10 hours)		

5.1	Recurrent neural networks – Computational graphs, RNN design	1 hour
5.2	Encoder – decoder sequence to sequence architectures	1 hour
5.3	Deep recurrent networks- Architecture	1 hour
5.4	Recursive neural networks	1 hour
5.5	Modern RNNs - LSTM	1 hour
5.6	Modern RNNs - LSTM	1 hour
5.7	GRU	1 hour
5.8	Practical use cases for RNNs.	1 hour
5.9	Case study - Natural Language Processing.	1 hour
5.10	Case study - Natural Language Processing.	1 hour