AIT 401	FOUNDATIONS OF DEEP LEARNING	CATEGORY	L	T	P	CREDIT
		PCC	2	1	0	3

Preamble: Study of this course provides the learners an overview of the concepts and algorithms involved in deep learning. The course covers the basic concepts in neural networks, deep learning, optimization techniques, regularization techniques, convolutional neural networks, recurrent neural networks, autoencoders, generative models. The students will be able to implement deep learning algorithms to solve real-world problems.

Prerequisite: Machine learning concepts

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

CO 1	Illustrate the basic concepts of neural networks, deep learning and its practical issues **(Cognitive Knowledge Level : Apply)**
CO 2	Outline the standard regularization and optimization techniques for the effective training of deep neural networks. **(Cognitive Knowledge Level: Understand)**
CO 3	Build convolutional Neural Network (CNN) models for different use cases. **(Cognitive Knowledge Level: Apply)**
CO 4	Apply the concepts of Recurrent Neural Network (RNN), Long Short Term Memory(LSTM), Gated Recurrent Unit (GRU)**. (Cognitive Knowledge Level: Apply)**
CO 5	Explain the concepts of auto encoder, generative models **(Cognitive Knowledge Level: Understand)**

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO6	PO 7	PO 8	PO 9	PO 10	PO1 1	PO 12
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 Finance
PO6	The Engineer and Society	**PO12**	Life long learning

Assessment Pattern

Bloom's Category	Continuous Assessment Tests		End Semester Examination Marks
	Test1 (percentage)	Test2 (percentage)	
Remember	20	20	20
Understand	40	40	40
Apply	40	40	40
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 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 maximum 2 sub-divisions and carries 14 marks.

Syllabus

Module 1: Introduction to Neural Networks and Deep learning

Introduction, The Basic Architecture of Neural Networks - Single Computational Layer: The Perceptron, Multilayer Neural Networks. Activation functions – Sign, Sigmoid, Tanh, ReLU, leaky ReLU, Hard Tanh, Softmax. Loss function. Training a Neural Network with Backpropagation. Practical issues in neural network training. Overfitting, Underfitting, Hyper parameters and Validation sets, Estimators -Bias and Variance.Introduction to deep learning, Deep feed forward network.

Module 2: Training deep models

Introduction, setup and initialization- Kaiming, Xavier weight intializations, Vanishing and exploding gradient problems, Optimization techniques - Gradient Descent (GD), Stochastic GD, GD with momentum, GD with Nesterov momentum, AdaGrad, RMSProp, Adam., Regularization Techniques - L1 and L2 regularization, Early stopping, Dataset augmentation, Parameter tying and sharing, Ensemble methods, Dropout, Batch normalization.

Module 3: Convolutional Neural Networks

Convolutional Neural Networks –Architecture, Convolution operation, Motivation, pooling .Variants of convolution functions, Structured outputs, Data types, Efficient convolution algorithms, Applications of Convolutional Networks, Pre-trained convolutional Architectures : AlexNet, ZFNet, VGGnet-19, ResNet-50.

Module 4: Recurrent Neural Networks

Recurrent neural networks – Computational graphs. RNN design. Encoder – decoder sequence to sequence architectures. Language modeling example of RNN. Deep recurrent networks. Recursive neural networks. Challenges of training Recurrent Networks. Gated RNNs LSTM and GRU.

Case study: BERT, Social Media Sentiment Analysis.

Module 5: Auto-encoders and Generative models.

Autoencoders, *Variational Auto-Encoder*-under complete Auto-encoder, stochastic encoder, denoising encoder, Applications of Autoencoders. Generative models - Boltzmann machines, Deep Belief Networks, Generative Adversarial Networks.

Reference Books

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. Deep Learning, Core Conceps, Methods and Applications- M Gopal, Pearson Education
4. Fundamentals of Deep Learning: Designing Next-Generation Machine Intelligence Algorithms (1st. ed.). Nikhil Buduma and Nicholas Locascio. 2017. O'Reilly Media, Inc.

Sample Course Level Assessment Questions

Course Outcome 1 (CO1):

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.
3. Sketch the typical learning curves for the training and validation sets, for a setting where

overfitting occurs at some point. Assume that the training set and the validation set are of the same size.

Course Outcome 2 (CO2):

1. Explain how L2 regularization improves the performance of deep feed forward neural networks.
2. Explain how L1 regularization method leads to weight sparsity.
3. Derive update rules for parameters in the multi-layer neural network through the gradient descent.

Course Outcome 3(CO3):

1. Give two benefits of using convolutional layers instead of fully connected ones for visual tasks.
2. 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?
3. Weight sharing allows CNNs to deal with image data without using too many parameters. Does weight sharing increase the bias or the variance of a model?

Course Outcome 4 (CO4):

1. Illustrate the workings of the RNN with an example of a single sequence defined on a vocabulary of four words.
2. List the differences between LSTM and GRU
3. Show the steps involved in an LSTM to predict stock prices.Give one advantage of using an RNN rather than a convolutional network.

Course Outcome 5 (CO5):

1. Is an autoencoder for supervised learning or for unsupervised learning? Explain briefly.
2. List the difference between Boltzmann Machine and Deep Belief Network.

Model Question paper

QP CODE: PAGES:3

Reg No:_____________

Name :_____________

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

EIGHTH SEMESTER B.TECH DEGREE EXAMINATION, MONTH & YEAR

***Course Code:* AIT 401**

Course Name: Foundations of Deep Learning

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

PART A

Answer all Questions. Each question carries 3 Marks

1. Illustrate the limitation of a single layer perceptron with an example
2. Specify the advantages of ReLU over sigmoid activation function.
3. Derive weight updating rule in gradient descent when the error function is a) mean squared error b) cross entropy
4. List any three methods to prevent overfitting in neural networks
5. Illustrate the strengths and weaknesses of convolutional neural networks.
6. What happens if the stride of the convolutional layer increases? What can be the maximum stride? Justify your answer
7. List the differences between LSTM and GRU
8. How does a recursive neural network work?
9. List the difference between Boltzmann Machine and Deep Belief Network.
10. How does the variational auto-encoder(VAE) architecture allow it to generate new data points, compared to auto-encoder, which cannot generate new data points?

(10x3=30)

Part B

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

11.

a. Explain back propagation algorithm for neural network training. (9 marks)

b. "How does bias and variance trade-off affect machine learning algorithms? (5 marks)

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 marks)

b. Compare overfitting and underfitting. How it can affect model generalization ? (6 marks)

13.

a. 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 marks)

b. Describe the effect in bias and variance when a neural network is modified with more number of hidden units followed with dropout regularization. (6 marks)

OR

14.

a. Explain how L2 regularization improves the performance of deep feed forward neural networks. (7 marks)

b. Initializing the weights of a neural network with very small or large random numbers is not advisable. Justify. (7 marks)

15.

a. 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. (6 marks)

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?
(8 marks)

OR

16.

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

b. What is the motivation behind convolution neural networks? (4 marks)

17. a. If we have a recurrent neural network (RNN), we can view it as a different type of network by "unrolling it through time". Briefly explain what that means. (6 marks)

b. Explain the architecture of GRU. (8 marks)

OR

18.

a. The vanishing gradient problem is more pronounced in RNN than in traditional neural networks. Give reason. Discuss a solution for the problem. (7 marks)

b. Show the steps involved in an LSTM to predict stock prices.Give one advantage of using an RNN rather than a convolutional network. (7 marks)

19.

a. Generative Adversarial Networks(GANs) include a generator and a discriminator. Sketch a basic GAN using those elements, a source of real images, and a source of randomness. (10 marks)

b. The word "adversarial" in the acronym for GANs suggests a two-player game. What are the two players, and what are their respective goals? (4 marks)

OR

20.

a. Explain auto encoder with an example. (7 marks)

b. Explain Generative Adversarial Networks using suitable diagram. (7 marks)

(14X5=70)

Teaching Plan		
No	Topic	No. of Lectures (**36 Hours**)
1	**Module 1: Introduction to neural network and Deep Learning**	**7**
1.1	Introduction, The Basic Architecture of Neural Networks - Single Computational Layer: The Perceptron.	1 hour
1.2	Multilayer Neural Networks.	1 hour
1.3	Activation functions - Sigmoid, Tanh, ReLU, leaky ReLU, Hard Tanh, Softmax. Loss function.	1 hour
1.4	Training a Neural Network with Backpropagation.	1 hour
1.5	Practical issues in neural network training	1 hour
1.6	Overfitting, Underfitting, Hyper parameters, Validation sets	1 hour
1.7	Estimators -Bias and Variance, Introduction to deep learning, Deep feed forward network	1 hour
2	**Module 2: Training deep models**	**8**
2.1	Introduction, setup and initialization issues- Kaiming and Xavier weight initializations	1 hour
2.2	Vanishing and exploding gradient problems	1 hour
2.3	Concepts of optimization, Gradient Descent (GD)	1 hour
2.4	Stochastic GD, GD with momentum, GD with Nesterov momentum	1 hour
2.5	AdaGrad, RMSProp, Adam	1 hour
2.6	Concepts of Regularization, L1 and L2 regularization	1 hour
2.7	Early stopping, Dataset augmentation	1 hour

2.8	Parameter tying and sharing, Ensemble methods, Dropout, Batch Normalization	1 hour
3	**Module 3: Convolutional Neural Network**	**8**
3.1	Convolutional Neural Networks, Architecture	1 hour
3.2	Convolution operation	1 hour
3.3	Motivation, pooling	1 hour
3.4	Variants of convolution functions	1 hour
3.5	Structured outputs, Data types	1 hour
3.6	Efficient convolution algorithms	1 hour
3.7	Applications of Convolutional Networks	1 hour
3.8	Case Studies of Convolutional Architectures : AlexNet, ZFNet, VGGNet-19, ResNet-50	1 hour
4	**Module 4 : Recurrent Neural Network**	**7**
4.1	Recurrent neural networks – Computational graphs	1 hour
4.2	RNN design, Encoder – decoder sequence to sequence architectures	1 hour
4.3	Language modeling example of RNN	1 hour
4.4	Deep recurrent networks, Recursive neural networks, Challenges of training Recurrent Networks	1 hour
4.5	LSTM	1 hour
4.6	GRU	1 hour
4.7	Case Study- BERT, Sentiment Analysis	1 hour
5	**Module 5 : Autoencoders and Generative models**	**6**
5.1	Autoencoders	1 hour
5.2	VariationalAutoEncoder , Applications of Autoencoders	2 hour
5.3	Boltzmann machines,	1 hour
5.4	Deep Belief Networks,	1 hour

5.5	Generative Adversarial Networks.	1 hour