Birla Institute of Technology & Science, Pilani
Work Integrated Learning Programmes Division
****

Digital Learning Handout

Part A: Content Design

Course Title	Deep Neural Networks
Course No(s)	AIML * ZG511
Credit Units	4
Credit Model	3-1-0
Course Author	Prof. Seetha Parameswaran
Version No:	2
Date:	20/09/2025

Course Description:
Introduction to neural networks, approximation properties, back propagation, deep network training, regularization and optimization, convolution neural networks, recurrent neural networks, attention models, transformers, neural architecture search, federated learning, meta learning, applications in time series modelling and forecasting, online (incremental) learning.

Course Overview:
This course provides comprehensive training in deep neural networks, covering fundamental architectures including feedforward networks, CNNs, RNNs, and transformers. Students will master training techniques, optimization algorithms, and regularization methods while implementing end-to-end deep learning solutions for computer vision, natural language processing, and time series applications.

Course Objectives

No	Course Objective
CO1	Apply deep learning architectures (feedforward, convolutional, recurrent, and transformer networks) to solve complex problems in computer vision, natural language processing, and time series analysis.
CO2	Analyze the impact of hyperparameters, optimization algorithms, and regularization techniques on deep learning model performance, training dynamics, and generalization capabilities.
CO3	Evaluate different deep learning architectures and techniques for specific problem domains by considering computational complexity, data requirements, model interpretability, and performance trade-offs.
CO4	Create innovative deep learning solutions that integrate multiple architectures and advanced techniques including attention mechanisms and transformers to address real-world challenges.

Text Book(s):

T1	Zhang, A., Lipton, Z. C., Li, M., & Smola, A. J. (2023). Dive into deep learning. Cambridge University Press.

Reference Book(s) & other resources:

R1	Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). Deep learning (Vol. 1, No. 2). Cambridge: MIT press.
R2	Eugene, C. (2019). Introduction to deep learning. MIT Press
R3	Chollet, F., & Chollet, F. (2021). Deep learning with Python. Simon and Schuster.
R4	Jurafsky, D., & Martin, J. H. (Third Edition) Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition.

Learning Outcomes: Students will be able to

LO1	Implement various deep neural network architectures (feedforward, convolutional, recurrent, and transformer networks) and optimize them to solve complex problems in image recognition, natural language processing, and time series forecasting.
LO2	Analyze the impact of hyperparameters, optimization algorithms (SGD, Adam, RMSProp), and regularization techniques (dropout, batch normalization, weight decay) on model training, convergence, and generalization performance.
LO3	Evaluate the effectiveness and limitations of different deep learning architectures for specific tasks by comparing computational efficiency, accuracy, data requirements, and interpretability trade-offs.
LO4	Design and create advanced deep learning systems that incorporate attention mechanisms, multi-head attention, and transformer architectures to address challenging real-world problems across multiple domains.

Modular Content Structure

1. Introduction and Motivating Example
    - Deep learning - What, Why, where
    - Supervised learning Applications of deep learning
2. Artificial Neuron and Perceptron
    - Artificial Neuron vs Biological Neuron
    - Perceptron and Perceptron learning algorithm
3. Linear Neural Network for Regression
    - Single neuron in output layer for regression
4. Linear Neural Network for Classification
    - Single neuron in output layer for classification
    - Multiple neurons in output layer for classification
5. Deep Feedforward Neural Networks or Multi Layer Perceptrons
    - Non linearity and hidden layers

- Forward and backward propagation
- Impact of depth and width in DFNN

6. Convolutional Neural Networks
   - Convolution, Padding and stride, Channels, Pooling
   - Design of CNN
   - AlexNet, VGGNet, Inception Net, ResNet
   - Transfer Learning
7. Recurrent Neural Networks
   - Back-propagation through time
   - GRU, LSTM, BiLSTM
8. Attention Mechanism
9. Transformer
10. Optimization of Deep models
11. Regularization for Deep models

Part B: Learning Plan

Contact Session	List of Topic Title	Sub-Topics	Reference
1	**Introduction and Motivating Example**	• Deep Learning - definition • Understanding how deep learning works • Why Deep Learning? • Motivating Example of Wake word detection • Key components - Data, Models, Objective function, Optimization Algorithms • Supervised learning examples • Interacting with an Environment	T1 - Ch 1 R3 - Ch 1.1.1, 1.1.2. 1.1.5
2	**Artificial Neuron and Perceptron**	• Biological neuron • Artificial Neuron • Biological neuron vs artificial neuron • Connectionism model • Perceptron • Perceptron for NOT, AND, OR gates • Perceptron learning algorithm for logic gates • Perceptron for linearly separable data • XOR Problem - demonstrate perceptron fails to learn - non linearly separable data • Code demonstrating AND, OR, XOR implementation from scratch	http://mlsp.cs.cmu.edu/people/rsingh/docs/Chapter1_Introduction.pdf
3	**Linear Neural Networks for Regression**	• Single neuron for regression - how much or how many • Data, linear model with single neuron and squared loss function for single neuron for regression (no hidden layers)	T1 – Ch 3 T1 - Ch 12.5 (Mini batch SGD)

		• Training using mini-batch stochastic gradient descent algorithm • Prediction or inference • Eg: Auto MPG (UCI) prediction with a single neuron - code implementation from scratch	
4	**Linear Neural Networks for Classification**	• Single neuron for classification - which category • Data, linear model with single neuron, sigmoid activation function, and binary cross entropy loss for single neuron for classification (no hidden layers) • Training using mini-batch stochastic gradient descent algorithm • Prediction or inference • Eg: Breast Cancer Wisconsin (Diagnostic) (UCI) prediction with a single neuron - code implementation from scratch • Multi-class classification - which category • Data, linear model with multiple output neuron, softmax activation function, and cross entropy loss for multiple neurons for classification (no hidden layers) • Training using mini-batch stochastic gradient descent algorithm • Prediction or inference • Eg: Iris (UCI) prediction with a multiple neuron - code implementation from scratch	T1 – Ch 4 T1 - Ch 12.5 (Mini batch SGD)
5	**Deep Feedforward Neural Networks (DFNN) or Multi Layer Perceptron (MLP) for Classification**	• MLP as solution for XOR • Hidden layers and Non-linearity • Forward Propagation (use vectorization and algorithm for forward pass) • Define loss functions and compute the error • Backward Propagation (use vectorization and algorithm for backprop) • Weight updation using gradients (use vectorization and algorithm for backprop) • Impact of depth and width in DFNN • Code implementation from scratch	T1 – Ch 5 R4 - Ch 7.3, 7.5
6	**Convolutional Neural Networks**	• Image Data • Invariance, Locality and Channels • Convolutions operation • Learning a Kernel	T1 – Ch 7 R3 - 2.2.11

		• Feature map and receptive field • Padding and Stride • 1x1 convolution • Pooling • Multiple channels • Putting it all together in LeNet • Code implementation	
7	**Deep Convolutional Neural Networks**	• AlexNet - Representation Learning • Networks Using Blocks (VGG) - VGG Blocks • Network in Network (NiN) - NiN Blocks and NiN Model • GoogLeNet - Inception Blocks • Residual Networks (ResNet) - Residual Blocks	T1 – Ch 8
8	**Convolutional Networks**	• Transfer Learning • Fine tuning for transfer learning • Review	Class notes
9	**Recurrent Neural Networks**	• Raw Text into Sequence Data • Recurrent connection • Recurrent Neural Networks with Hidden States • Backpropagation Through Time • RNNs for NLP tasks • Encoder–Decoder Architecture • Teacher Forcing • Loss Function with Masking • Training and Prediction	T1 - Ch 9 R4 – Ch 9 R3 - 2.2.10
10	**Deep Recurrent Neural Networks**	• Long Short-Term Memory (LSTM) • Gated Memory Cell • Gated Recurrent Units (GRU) • Reset Gate and Update Gate, Candidate Hidden State • Stacked RNN architectures • Bidirectional Recurrent Neural Networks • Code implementation	T1 – Ch 10 R4 – Ch 9
11	**Attention Mechanism**	• Queries, Keys, and Values • Attention Pooling by Similarity • Attention Pooling via Nadaraya–Watson Regression • Attention Scoring Functions	T1 – Ch11 R4 - Ch 9

		• Dot Product Attention • Convenience Functions • Scaled Dot Product Attention • Additive Attention • Bahdanau Attention Mechanism • Multi-Head Attention • Self-Attention • Positional Encoding • Code implementation	
12	**Transformer**	• Transformer architecture • Model, Positionwise Feed-Forward Networks, Residual Connection and Layer Normalization • Encoder and Decoder • Transformer block • Residual view for transformer	T1 – Ch 11 R4 – Ch10
13	**Transformer**	• Transformers for Vision • Model, Patch Embedding, Vision Transformer Encoder, Training and Evaluation • Large-Scale Pretraining with Transformers • Encoder-Only, Encoder–Decoder, Decoder-Only • Scalability	T1 – Ch 11 R4 - Ch 10.7
14	**Optimization of Deep models**	• Goal of Optimization • Optimization Challenges in Deep Learning • Gradient Descent • Stochastic Gradient Descent • Minibatch Stochastic Gradient Descent • Momentum • Adagrad and Algorithm • RMSProp and Algorithm • Adadelta and Algorithm • Adam and Algorithm • Implementation of algorithm from scratch and comparison of algorithms	T1 – Ch 12
15	**Regularization for Deep models**	• Generalization for regression • Training Error and Generalization Error • Underfitting or Overfitting • Model Selection • Weight Decay and Norms	T1 – Ch 3.6, 3.7 T1 - Ch 4.6, 4.7 T1 - Ch 5.5, 5.6 T1 - Ch 8.5 T1 - Ch 11.7

		• Generalization in Classification • Environment and Distribution Shift • Generalization in Deep Learning • Dropout • Batch Normalization • Layer Normalization • Code implementation	
16	**Review**	Review of contact session 1 to 15	

Experiential Learning Components:

1. Lab work: 10
2. Project work: 0
3. Case Study: 0
4. Simulation: 0
5. Work Integrated Learning Assignment- 2 Assignments
6. Design work/ Field work: 0

Objective of Experiential Learning Component:

Hands on sessions on implementation of

a) Linear neural networks for regression and classification
b) Deep feedforward neural networks (MLP) with backpropagation
c) Convolutional Neural Networks (CNN) for image recognition
d) Deep CNN architectures (AlexNet, VGG, ResNet) and transfer learning
e) Recurrent Neural Networks (RNN, LSTM, GRU) for sequence modeling
f) Attention mechanisms and multi-head attention
g) Transformer architectures for NLP and vision tasks
h) Optimization algorithms (SGD, Momentum, Adam, RMSProp)
i) Regularization techniques (dropout, batch normalization, weight decay)

Scope of Experiential Learning Component:

Programming language: Python
Tools and libraries: Jupyter, Numpy, Scipy, Pandas, ScikitLearn, Matplotlib, Seaborn, Tensorflow, Pytorch, Keras
Development Environment: - Python 3.8+ - TensorFlow 2.x / PyTorch 2.0+

Lab Infrastructure:

WILP Virtual lab

BITS Pilani
Pilani | Dubai | Goa | Hyderabad

List of Experiments:

Lab No	Lab Objective	Session Reference
1	Building linear neural network for regression: Auto MPG prediction with single neuron implementation from scratch	3
2	Implementing linear neural network for binary classification: Breast Cancer Wisconsin prediction with single neuron from scratch	4
3	Building multi-class classification with multiple output neurons: Iris dataset prediction from scratch	4
4	Implementing Deep Feedforward Neural Network (MLP) with forward and backward propagation from scratch	5
5	Building Convolutional Neural Network (CNN) with LeNet architecture for image classification	6
6	Implementing deep CNN architectures (AlexNet/VGG/ResNet) and transfer learning for image recognition	7, 8
7	Building Recurrent Neural Networks (RNN, LSTM, GRU) for sequence modeling and text processing	9, 10
8	Implementing attention mechanisms and Transformer architecture for NLP tasks	11, 12, 13
9	Comparing optimization algorithms (SGD, Momentum, Adam, RMSProp, Adagrad, Adadelta) with implementation from scratch	14
10	Implementing and comparing regularization techniques (dropout, batch normalization, layer normalization, weight decay)	15

Evaluation Scheme:

Legend: EC = Evaluation Component; AN = After Noon Session; FN = Fore Noon Session

Evaluation Component	Name (Quiz, Lab, Project, Mid-term exam, End semester exam, etc.)	Type (Open book, Closed book, Online, etc.)	Weight	Duration	Day, Date, Session, Time
EC – 1*	Quiz	Online	10%	1 week	To be announced
	Practice Labs Situated Learning assignment	Online	20 %	10 days	To be announced
EC - 2	Mid-Semester Test	Closed Book	30%	2 hours	To be announced
EC - 3	Comprehensive Exam	Open Book	40%	2 ½ Hours	To be announced

EC1* (20% - 30%): Quiz (optional): 5-10 %, Lab Assignment/Assignment: 20% - 30%

Syllabus for Mid-Semester Test (Closed Book): Topics in Contact session: 1 to 8

Syllabus for Comprehensive Exam (Open Book): All topics

Important Links and Information:

eLearn Portal: https://elearn.bits-pilani.ac.in

Students must visit the eLearn portal regularly and stay updated with the latest announcements and deadlines.

Contact Sessions: Students should attend the online lectures as per the schedule provided on the eLearn portal.

Evaluation Guidelines:

1. EC-1 consists of either two Assignments or three Quizzes. Students will attempt them through the course pages on the eLearn portal. Announcements will be made on the portal in a timely manner.
2. For Closed Book tests: No books or reference material of any kind will be permitted.
3. For Open Book exams: "open book" means text/ reference books (publisher copy only) and does not include any other learning material. No other learning material will be permitted during the open book examinations. For Detailed Guidelines refer to the attached document.
   EC3 Guidelines
4. If a student is unable to appear for the Regular Test/Exam due to genuine exigencies, the student should follow the procedure to apply for the Make-Up Test/Exam, which will be made available on the eLearn portal. The Make-Up Test/Exam will be conducted only at selected exam centres on the dates to be announced later.

It shall be the responsibility of the individual student to be regular in maintaining the self-study schedule as given in the course handout, attend the online lectures, and take all the prescribed evaluation components such as Assignments/Quizzes, Mid-Semester Tests and Comprehensive Exams according to the evaluation scheme provided in the handout.

****