

**VASIREDDY VENKATADRI INSTITUTE OF TECHNOLOGY NAMBUR-522508 ANDHRA PRADESH, INDIA**

**YEAR :** III B.Tech **SEMESTER:** II

**COURSE NAME:** DEEP LEARNING

**COURSE CODE:** XXXXXXXX

**BRANCH:** CSM

**PREREQUISITE:** DIGITAL IMAGE PROCESSING, LINEAR ALGEBRA, CALCULUS, PROBABILITY THEORY

**COURSE OBJECTIVE:** To create a strong foundation on deep learning architectures & algorithms for building models in computer vision, natural language processing related application areas.

**COURSE OUTCOMES:** Students will be able to:

|  |  |  |  |
| --- | --- | --- | --- |
| **SN** | **OUTCOME** | **Cognitive Levels as per Bloom’s Taxonomy** | **Weightage (%)** |
| CO1 | Demonstrate the building of fully connected feedforward neural networks | L1, L2 | 20 |
| CO2 | Perform optimized training of the feedforward neural networks | L1, L2, L3, L4 | 20 |
| CO3 | Design convolutional neural networks for solving basic computer vision problems | L1, L2, L3, L4 | 20 |
| CO4 | Apply effective training techniques | L1, L2, L3, L4 | 20 |
| CO5 | Design RNNs for NLP | L1, L2, L3, L4 | 20 |

**WEIGHTAGE OF BLOOM’S LEGENDS & PERCENTAGE OF QUESTIONS IN EXAMINATIONS:**

L1 (Remembering) = 30 - 40%, L2 (Understanding) = 30 - 40%,

L3 (Applying) = 10 - 20 %, L4 (Analysing) = 10 - 20%,

Easy (%) = 15%-20%, Average (%) = 60% - 70%, Difficult (%) = 15% - 20%

TOTAL = L1 + L2 + L3 + L4 = 100% (on an average about 2 minutes per mark)

**Note:** This specification weightage in above shall be treated as a general guideline for students, teachers and paper setters. The actual distribution of marks in the question paper may vary slightly.

**DETAILED SYLLABUS:**

**UNIT-1: Deep learning basics:** Introduction, the perceptron, Overfitting and generalization, linear perceptron, learning XOR function with non-linear functions, feedforward neural networks, types of activation functions, types of loss functions, Back-Propagation.

**UNIT-II: Optimization:** Challenges in neural network optimization, Regularization, Gradient Descent, Stochastic Gradient Descent, Momentum Optimizer, AdaGrad, RMSProp, Adam, Batch normalization.

**UNIT-III: Deep Learning for Computer Vision:** Building blocks of CNN, Local receptive fields, Shared weights and bias, stride, Pooling layers, Max-pooling, Average pooling,  CNN for image classification – AlexNet, VGG, GoogleNet, ResNet architectures. CNN for segmentation – Unet.

**UNIT-IV: Effective training of Deep Neural Networks:** Early stopping, Dropout, Instance Normalization, Group Normalization, Transfer Learning, Data Augmentation.

**UNIT-V: Deep Learning for Natural Language Processing:** Computational representation of language, one-hot representation of words, word vectors – the skip-gram word2vec model, The CBOW word2vec model, word vector arithmetic, RNN, LSTM.

**TEXT BOOKS:**

1. Deep Learning- Ian Goodfellow, Yoshua Benjio, Aaron Courville, The MIT Press
2. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2 Edition - Aurélien Géron, O'Reilly Media, Inc. ISBN: 9781492032649
3. Pattern Classification- Richard O. Duda, Peter E. Hart, David G. Stork, John Wiley & Sons Inc.

**REFERENCE BOOKS:**

1. Theodoridis, S. and Koutroumbas, K. Pattern Recognition. Edition 4. Academic Press, 2008.
2. Russell, S. and Norvig, N. Artificial Intelligence: A Modern Approach. Prentice Hall Series in Artificial Intelligence. 2003.
3. Bishop, C. M. Neural Networks for Pattern Recognition. Oxford University Press. 1995.
4. Hastie, T., Tibshirani, R. and Friedman, J. The Elements of Statistical Learning. Springer. 2001.
5. Koller, D. and Friedman, N. Probabilistic Graphical Models. MIT Press. 2009.

**ONLINE REFERENCES:**

1. NPTEL Lecture material - Lecture Series on Deep Learning by Prof. P. K. Biswas, Department of Electrical & Electronic Communication Engineering, IIT Kharagpur.

[https://onlinecourses.nptel.ac.in/noc22\_cs22/preview#:~:text=Week%201%3A%20Introduction%20to%20Deep,Multilayer%20Perceptron%2C%20Back%20Propagation%20Learning](about:blank#:~:text=Week%201%3A%20Introduction%20to%20Deep,Multilayer%20Perceptron%2C%20Back%20Propagation%20Learning)

**MICRO-SYLLABUS:**

|  |  |  |
| --- | --- | --- |
| **Unit** | **Module** | **Micro Content** |
| **UNIT I** | Introduction | Feature space, linearly separable classes |
| Overfitting and generalization | Capacity of the model selected and complexity of the task |
| The perceptron | How it learns linear functions, and fails to learn XOR function |
| Feedforward Neural Networks | How the nonlinear units enable the learning of any nonlinear function |
| Types of activation functions | Types of loss functions used in neural networks, their derivatives and use cases. |
| Types of loss functions | Regression loss functions: MSE, MAE, Huber loss function, classification loss functions: binary cross-entropy, categorical cross-entropy |
| Back-propagation | Propagation errors form output layer back to the input layer via hidden layers for updating parameters using gradient descent algorithms, |
| Examples of back propagation algorithm |
| **Unit** | **Module** | **Micro Content** |
|  | Introduction | Understanding loss landscape |
| **UNIT II** | Challenges in neural network optimization |
| Regularization | L2 and L1 regularization |
| Gradient Optimization and its variants | Gradient descent, and SGD |
| Momentum optimizer |
| AdaGrad |
| RMSProp |
| Adam |
| Batch normalization |
| **Unit** | **Module** | **Micro Content** |
| **UNIT III** | Building blocks of CNN | 1D and 2D Convolutional layers, stride, and Parameter calculations. |
| Local receptive fields, shared weights and biases |
| Pooling layers: Max-pooling, Average pooling |
| CNN for image classification | AlexNet, VGG, GoogleNet, ResNet architectures |
| CNN for image segmentation | Unet |
| **Unit** | **Module** | **Micro Content** |
| **UNIT IV** | Early stopping | Avoid overfitting using early stopping |
| Feature Normalization | Instance normalization, group normalization |
| Dropout | Dropout as regularization |
| Transfer learning | Applying transfer learning for downstream tasks for faster training |
| Data augmentation | Importance of data augmentation, types of data augmentations with respect to image classification. |
| **Unit** | **Module** | **Micro Content** |
| **UNIT V** | Computational representation of language | Vector representation of words, one-hot encoding |
| word vectors | skip-gram word2vec model |
| The CBOW word2vec model |
| word vector arithmetic |
| RNN | RNN |
| LSTM | LSTM |

Code No :

**R20**

**III B. TECH II SEMESTER REGULAR EXAMINATION MODEL PAPER**

**DEEP LEARNING**

**(CSM BRANCH)**

**Time : 3 Hours Max. Marks : 70**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Note :** Answer **ONE** question from each unit **(5 × 14 = 70 Marks)**

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| UNIT-I | | | | CO | BL |
| 1. | a) | Explain the need for nonlinear activation functions in neural networks. How the ReLU activation function solves the XOR problem? | [7M] | CO1 | L2 |
| b) | What is the choice of output layer units in the task of multiclass classification and why? | [7M] | CO1 | L2 |
| (OR) | | | |  |  |
| 2. | a) | How a linear neuron is converted into non-linear neuron? Explain why it is required in practice. | [7M] | CO1 | L2 |
| b) | How the cost functions are defined for various machine learning task performed using deep neural networks? | [7M] | CO1 | L1 |
| UNIT-II | | | |  |  |
| 3. | a) | What is meant by the regularization of a deep learning model? What different strategies are used for regularization? | [7M] | CO2 | L3 |
| b) | Explain in detail the process of ridge regularization and how it reduces the generalization error. | [7M] | CO2 | L1 |
| (OR) | | | |  |  |
| 4. | a) | Distinguish Adagrad optimization from the basic gradient descent optimization. | [7M] | CO2 | L3 |
| b) | What is the chain rule of calculus and how is it useful in a backpropagation algorithm? | [7M] | CO2 | L1 |
| UNIT-III | | | |  |  |
| 5. | a) | How CNNs are different from traditional neural networks with all fully connected layers? | [7M] | CO3 | L2 |
| b) | What is the context between the depth and filter size in building the CNN models | [7M] | CO3 | L3 |
| (OR) | | | |  |  |
| 6. | a) | How important it is to have skip connections in the ResNet blocks? Explain. | [7M] | CO3 | L4 |
| b) | The input consists of 7 channels of size 16x16. Use convolution later. Find the number of weights in each of the configurations below.  (a) The output of the first layer consists of 8 feature maps, and the filters are of size 5x5. The convolution is done with stride 2 and zero padding is used.  (b) Now suppose we made this a fully connected layer, but the number of input and output units are kept the same as in the network described in the Part (a) above. | [7M] | CO3 | L4 |
| UNIT-IV | | | |  |  |
| 7. | a) | How the overfitting can be avoided using drop out strategy? | [7M] | CO4 | L1 |
| b) | In which cases transfer learning is useful and how transfer learning is carried out? | [7M] | CO4 | L3 |
| (OR) | | | |  |  |
| 8. | a) | What are different normalization techniques used in CNNs? Explain. | [7M] | CO4 | L1 |
| b) | List out various training strategies for the better performance of the neural network models. | [7M] | CO4 | L3 |
| UNIT-V | | | |  |  |
| 9. | a) | What are different computational representations of a language? Explain in detail. | [7M] | CO5 | L1 |
| b) | With a suitable example explain the process of word2vec model. | [7M] | CO5 | L2 |
| (OR) | | | |  |  |
| 10. | a) | Draw a depth 1 simple recurrent network for language translation with inputs (x1, x2, ...., xn) and the outputs (y1, y2, …., yn). Explain how it works. | [7M] | CO5 | L2 |
| b) | In the LSTM model, explain exactly how the cell state is updated from Ct-1 to Ct, using the previous stateht-1 and the current input xt. | [7M] | CO5 | L3 |

**\* \* \* \* \***

**THE ABOVE MODEL PAPER ATTAINMENTS OF BLOOM’S TEXONOMY AS FOLLOWS**

**L1: 6\*7 = 42 = 30%**

**L2: 6\*7 = 42 = 30%**

**L3: 6\*7 = 35 = 30%**

**L4: 2\*7 = 14 = 10%**

SIGNATURES OF

COURSE COORDINATER MODULE COORDINATER HOD