BIRLA INSTITUTE OF TECHNOLOGY, MESRA, RANCHI
(END SEMESTER EXAMINATION)

CLASS: B.TECH **SEMESTER : VII**
BRANCH: COMPUTER SCIENCE AND ENGINEERING **SESSION : MO/2025**

SUBJECT: CS437 DEEP LEARNING

TIME: 3 HOURS **FULL MARKS: 50**

INSTRUCTIONS:
1. The question paper contains 5 questions each of 10 marks and total 50 marks.
2. Attempt all questions.
3. The missing data, if any, may be assumed suitably.
4. Before attempting the question paper, be sure that you have got the correct question paper.
5. Tables/Data hand book/Graph paper etc. to be supplied to the candidates in the examination hall.

			CO	BL
Q.1(a)	Define the Bayesian statistics theorem. Distinguish between the Maximum Likelihood Estimation (MLE) and Maximum A Posterior (MAP) for point estimation.	[2+3]	1	1, 4
Q.1(b)	A model achieves 98% accuracy on the training set but only 72% on the validation set. Diagnose whether the model is underfitting or overfitting. Describe three concrete steps you would take to correct the issue, and justify each step in terms of expected impact on generalization.	[5]	1	3, 5
Q.2(a)	Explain how mini-batch SGD differs from full-batch and pure stochastic gradient descent. Discuss why mini-batching is generally the most practical choice for large-scale datasets.	[2+3]	3	2
Q.2(b)	Describe how denoising autoencoders learn robust representations compared to standard autoencoders. Explain your answer with a suitable diagram. Describe how tying the encoder and decoder weights acts as a regularization technique in autoencoders.	[3+2]	2, 3	2
Q.3(a)	Explain the importance of Principal Component Analysis (PCA) in machine learning. Compute principal component for the given dataset 3x2 D=[[4,0][3,5][2,4]]	[2+3]	1	3,4
Q.3(b)	Show that injecting small Gaussian noise $\varepsilon \sim N(0,\sigma^2)$ into the inputs of a neural network during training is equivalent to adding an L2 regularization term to the loss function.	[5]	3	3,4
Q.4(a)	Explain how weight sharing and local connectivity reduce the number of parameters in CNNs. Explain why weight sharing is generally *not* used between the generator and discriminator in standard GAN architectures.	[3+2]	4	2,4
Q.4(b)	The cost function for a model is $J(w)=(w-3)^2+4$. Using gradient descent with learning rate $\alpha=0.1$ and initial weight $w_0=0$, compute the updated weights for two iterations.	[5]	3	3
Q.5(a)	Define and explain the concepts of **convolution**, **padding**, **stride**, and **pooling** in Convolutional Neural Networks (CNNs). Use neat diagrams to illustrate each operation clearly. Given an input feature map of size 5×5, a filter of size 3×3, stride = 1, and zero padding, compute the resulting output size using the standard convolution formula.	[3+2]	4	1,3
Q.5(b)	What is vanishing gradients problem in Recurrent Neural Networks (RNNs)? Describe one effective technique for mitigating the vanishing gradients problem. Explain the role of the forget gate in an LSTM.	[2+2+1]	4,5	2,3,4

:::::19/11/2025:::::E