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

CLASS: B.TECH **SEMESTER : VI**
BRANCH: AI/ML **SESSION : SP/2025**

SUBJECT: AI305 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)	Write out the equations that define the network in the given Figure 1 below. There should be three equations to compute the three hidden units from the inputs and two equations to compute the outputs from the hidden units.(Assume biases and weights correspondingly assuming RELU function for hidden units).	[5]	1	III

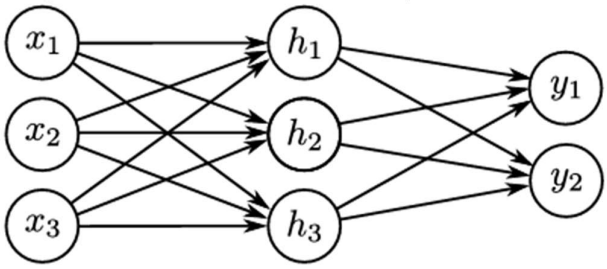

Figure 1. Shallow network.

Q.1(b)	In Figure 1 if x_1=1, x_2=2 , x_3=-1 , parameters from inputs to hidden units are, $(\theta_{10}, \theta_{20}, \theta_{30}, \theta_{11}, \theta_{12}, \theta_{13}, \theta_{21}, \theta_{22}, \theta_{23}, \theta_{31}, \theta_{32}, \theta_{33}) = (0,1,-2,1,-1,2,0,1,-1,2,0,1)$ and from hidden to output units are, $\emptyset = (\varphi_{10}, \varphi_{20}, \varphi_{11}, \varphi_{12}, \varphi_{13}, \varphi_{21}, \varphi_{22}, \varphi_{23}) =$(1,0,1,2,-1,0,-1,1). find the outputs y_1 and y_2 .(Assuming RELL for each hidden units).	[5]	1	III
Q.2(a)	Write out the equations for a deep neural network that takes D_i = 5 inputs, D_o = 4 outputs and has three hidden layers of sizes D_1 = 20, D_2 = 10, and D_3 = 7, respectively. What are the sizes of each weight matrix Ω. and bias vector ß.?	[5]	3	II
Q.2(b)	Derive the loss function for multiclass classification problems.	[5]	2	I
Q.3(a)	Illustrate the concept of Parameter initialization and explain the He initialization.	[5]	3	II
Q.3(b)	Explain the Gradient descent and Stochastic gradient descent method. Differentiate the difference between them with example.	[5]	3	I
Q.4(a)	Show that the weight decay parameter update with decay rate λ: $$\phi = (1-\lambda)\phi - \alpha\frac{dL}{d\phi},$$ on the original loss function *L[ϕ]* is equivalent to a standard gradient update using *L2* regularization so that the modified loss function $\widetilde{L[\phi]}$ is: $$\widetilde{L[\phi]} = L[\phi] + \frac{\lambda}{2\alpha}\sum_k \phi_k^2$$ where ϕ are the parameters, and *a* is the learning rate.	[5]	4	V
Q.4(b)	What are the sources of errors?. Suggest and elaborate the methods to reduce the errors.	[5]	4	I
Q.5(a)	Explain and demonstrate the notion of invariance and equivariance using an example.	[5]	5	II
Q.5(b)	Explain the following with example: (i) Padding (ii) Kernel size (iii) Stride (iv) Dilation. (v) Parameter sharing in convolutional layers	[5]	5	III

::::::29/04/2025::::::M