U.S.N.										

B.M.S. College of Engineering, Bengaluru-560019

Autonomous Institute Affiliated to VTU

June 2025 Semester End Main Examinations

Programme: B.E. **Semester: VI**

Branch: Artificial Intelligence and Machine Learning **Duration: 3 hrs.**

Course Code: 24AM6PCAGA **Max Marks: 100**

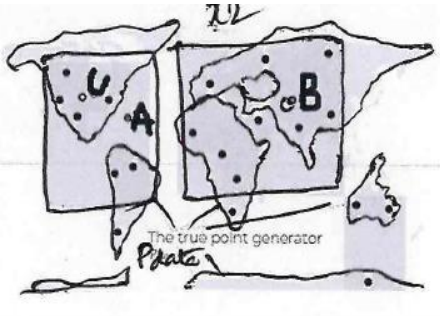
Course: Autoencoders and Generative AI

Instructions: 1. Answer any FIVE full questions, choosing one full question from each unit.
2. Missing data, if any, may be suitably assumed.

			UNIT - I	*CO*	*PO*	Marks
1	a)		Explain the key features of undercomplete autoencoders.	*C03*	*PO1*	**06**
	b)		Differentiate between Autoencoders and Variational Autoencoders.	*C02*	*PO2*	**06**
	c)		Consider a latent variable model with observed variable x and latent variable z. Let the generative model be $p_\theta(x,z)$ and the variational posterior approximation be $q_\phi(z\|x)$. i. Starting from log $p_\theta(x)$, derive the Evidence Lower Bound (ELBO) using the KL divergence. ii. Explain why maximizing the ELBO is equivalent to minimizing the KL divergence between the approximate posterior $q_\phi(z\|x)$ and the true posterior $p_\theta(z\|x)$.	*CO1*	*PO3*	**08**
			OR			
2	a)		How does Predictive Loss Decomposition use the loss for training? Explain.	*CO2*	*PO1*	**05**
	b)		Given a regularized autoencoder with a sigmoid activation function Input Vector X = [0.5, 0.9] Weight matrix (Input to Hidden): $W_1 = \begin{bmatrix} 0.4 & -0.2 \\ 0.1 & 0.3 \end{bmatrix}$ Bias Vector (Hidden Layer): b_1 = [0.0, 0.1] Weight matrix (Hidden to Output): $W_2 = \begin{bmatrix} 0.2 & 0.5 \\ -0.3 & 0.4 \end{bmatrix}$ Bias Vector (Output Layer): b_2 = [0.0, 0.0] L2 Regularization Parameter: λ =0.01 Compute the following: i. The reconstruction loss (Mean Squared Error). ii. The L2 regularization term. iii. The total loss.	*CO1*	*PO3*	**10**
	c)		Elaborate on Sparse Autoencoder with the associated loss function.	*CO3*	*PO1*	**05**

			UNIT - II			
	3	a)	A Hopfield network consists of 3 neurons with the following parameters: States s=[1,-1,1], Bias terms b=[0.2,-0.1,0.0] Weights W= $\begin{bmatrix} 0 & 0.5 & -0.4 \\ 0.5 & 0 & 0.3 \\ -0.4 & 0.3 & 0 \end{bmatrix}$ Calculate the total energy E of the network.	*CO1*	*PO3*	**06**
		b)	Write contrastive divergence algorithm and explain its working.	*CO2*	*PO1*	**08**
		c)	Explain how are Stacked Restricted Boltzmann Machines (RBMs) trained?	*CO3*	*PO3*	**06**
			OR			
	4	a)	A Machine Learning engineer is working on sentiment classification for product reviews in a target domain (e.g., electronics). However, only 500 labelled reviews are available in the target domain. To improve performance, the engineer decides to use Instance-Based Transfer Learning by borrowing data from a source domain (e.g., book reviews), which contains 10,000 labelled reviews. To address domain differences, instance re-weighting is applied: • The classifier is trained on all 10,000 source domain reviews and 500 target domain reviews. • Each source instance is given a weight of 0.3, and each target instance a weight of 1.0 during training. • The accuracy of the model trained without weighting is 68%. • The accuracy of the model trained with instance weighting improves to 75%. For the given scenario, answer the following: i. What is the purpose of instance re-weighting in this context? ii. Compute the effective weighted sample size used in training. iii. Calculate the relative improvement in accuracy due to re-weighting. iv. Name one challenge that can arise in instance-based transfer learning and explain briefly.	*CO1*	*PO3*	**08**
		b)	Write Greedy layer wise unsupervised pretraining protocol algorithm.	*CO2*	*PO1*	**06**
		c)	Differentiate between the 2 types of samples handled by Boltzmann machine while learning the weights.	*CO3*	*PO2*	**06**
			UNIT - III			
	5	a)	A Variational Autoencoder (VAE) uses a latent variable $z \sim N(\mu,\sigma^2)$ sampled from a Gaussian distribution. Directly sampling z introduces a non-differentiable operation, which obstructs standard backpropagation during training.	*CO1*	*PO3*	**08**

			i. Explain the reparameterization trick and how it allows gradients to flow through stochastic nodes. Include the expression used. ii. Why is it important that the random variable used in reparameterization is independent of μ and σ ? iii. Given, μ=2.0, σ=0.5, and a standard normal sample z=−1.2, compute the reparametrized latent variable y. iv. Compute the partial derivatives ∂y/∂μ and ∂y/∂σ.			
	b)	Explain any 2 types of Full Directed Network Models in detail.	*CO2*	*PO2*	**06**
	c)	Elaborate on the Ancestral sampling technique.	*CO3*	*PO3*	**06**
		OR			
6	a)	A software firm uses Neural Autoregressive Distribution Estimator (NADE) model for density estimation of high-dimensional binary data. i. What is the benefit of using Permutation-Invariant NADE, and how is it achieved during training? ii. Explain how model ensembles using multiple variable orderings can improve the likelihood estimation of NADE. Include the mathematical expression used for ensemble prediction. iii. Given a Deep NADE model with Input dimension n=100 and Hidden layer size h=50, Estimate the computational cost for: A. Regular NADE B. Deep NADE (unconstrained)	*CO1*	*PO3*	**08**
	b)	Elaborate on the 2 joint training approaches used to train Deep Boltzmann Machines.	*CO2*	*PO1*	**07**
	c)	Write the difference between Gibb's sampling and Ancestral sampling.	*CO3*	*PO2*	**05**
		UNIT - IV			
7	a)	Each avatar is defined by a unique combination of 5 different eye shapes{Round, Almond, Hooded, Monolid, Upturned}, 4 different skin tones {Light, Tan, Brown, Dark}, 3 different eyebrow styles {Straight, Arched, Curved}, 6 different mouth types {Smile, Open, Frown, Neutral, Pout, Wide}, 2 different background styles {SolidColor, Gradient}.Using the above statement , answer the following: i. How many unique avatars can be generated by choosing one option from each of the categories above? ii. If we fix the background style to SolidColor, how many unique avatars can be generated? iii. If avatars with Round eyes and Open mouth are considered invalid combinations, how many valid avatars remain?	*CO1*	*PO3*	**06**
	b)	Elaborate on any 2 real time applications based on the concept of Generative modeling.	*CO2*	*PO1*	**08**
	c)	How does Generative models differ from Discriminative models? Explain.	*CO3*	*PO3*	**06**

			OR			
	8	a)	In the given picture, rectangular-shaded region represents the model distribution p_{model}, and the map region shows the true data-generating distribution p_{data}. Points A (located in mid of the sea), B (located very close to existing point), and C (lies within the expected region) are samples drawn from p_{model}. Analyze the validity of these points based on the generative modeling framework rules provided and also conclude your observations with proper justification.  p_{model} wll be impressive if: **Rule 1:** It can generate examples that appear to have been drawn from p_{data}. **Rule 2:** It can generate examples that are suitably different from the observations in X.	*CO2*	*PO2*	**10**
		b)	List and explain any 5 benefits of Generative AI in detail.	*CO3*	*PO3*	**10**
			UNIT - V			
	9	a)	A Generative Adversarial Network is trained using a mini-batch of 4 samples. The discriminator gives the following output probabilities. Find the discriminator loss, generator loss and determine which is the best model with proper reasoning. **Sample Type** | **Input** | **Discriminator Output** Real | X_1 | 0.91 Real | X_2 | 0.87 Fake | $G(Z_1)$ | 0.21 Fake | $G(Z_2)$ | 0.18	*CO1*	*PO3*	**10**
		b)	Illustrate CycleGAN along with its objective, loss function and related numerical examples.	*CO2*	*PO3*	**10**
			OR			
	10	a)	Consider a Pix2Pix model where, the discriminator's output for the generated image G(x) is D(G(x)) =0.6, The ground truth image pixels are given by y= [0.2,0.4,0.6], The generated image pixels are G(x)=[0.3,0.5,0.7], The L1 loss weight parameter is λ=50. Answer the following: i. The adversarial loss for the generator based on D(G(x)). ii. The L1 loss between the ground truth image y and the generated image G(x). iii. Total generator loss.	*CO1*	*PO3*	**10**
		b)	Using suitable mathematical relations and numerical example, elaborate on the adversarial Loss of a Vanilla GAN.	*CO2*	*PO2*	**10**

******