Roll No: | | | | | | | | | | | |

PRESIDENCY UNIVERSITY

GAIN MORE KNOWLEDGE
REACH GREATER HEIGHTS

Presidency University Act, 2013 of the Karnataka Act No. 41 of 2013 | Established under Section 2(f) of UGC Act, 1956
Approved by AICTE, New Delhi | Approved By BCI
Bengaluru

Mid Term Examinations - October 2026

Date: 05/10/2026 **Time:** 11:45 AM - 01:15 PM

Course Code: CSG2501 **Course Name:** Gen AI for Data Science

Semester: Seventh Semester **Max. Marks:** 50 **Weightage:** 50%

CO - Levels	CO1	CO2
Marks	46	44

ANSWER KEY

PART A: Answer Following Questions. 10 M

Q.N.	Questions	M	CO	BT
1	Answer Traditional machine learning generally focuses on prediction, classification or decision-making from input data. Generative AI learns patterns in existing data and generates new data such as text, images or synthetic records.	2	CO1	BT2
2	Answer Consider email processing: **Discriminative model:** Predicts whether an email is spam or not spam. **Generative model:** Learns the characteristics of emails and can generate new email-like samples. Thus, discriminative models focus on prediction, while generative models focus on learning and generating data.	2	CO1	BT2
3	Probability distributions can be classified as: 1. **Discrete distributions** – used for discrete random variables. Example: Bernoulli distribution. 2. **Continuous distributions** – used for continuous random variables. Example: Gaussian distribution.	2	CO1	BT2
4	The latent space is the compact representation learned by the VAE. It captures important characteristics of the input data and allows the decoder to generate new samples. x→Encoder→z→Decoder	2	CO2	BT3
5	The two principal VAE loss components are: L=Lreconstruction+LKLL=L_{reconstruction}+L_{KL} The **reconstruction loss** measures how closely the reconstructed output matches the original input. For blurry reconstructions, reconstruction quality is directly related to the reconstruction-loss component.	2	CO2	BT3

PART B : Answer Any 1 Following Questions. 10 M

Q.N.	Questions	M	CO	BT

PTO (Please Turn Over)

6	Answer The major development stages can be summarized as: 1. **Rule-based AI** – relied on predefined rules. 2. **Machine Learning** – learned patterns from data. 3. **Deep Learning** – used multi-layer neural networks. 4. **Generative AI** – learns data patterns and generates new content. Advances in deep learning, large datasets and computational resources enabled the development of modern generative models.	10	CO1	BT2
7	Answer A **generative approach** learns the underlying characteristics or distribution of data and can generate new samples. A **discriminative approach** focuses on distinguishing between classes or predicting outputs. **Example:** For a spam classification problem: Email→Discriminative Model→Spam/Not SpamEmail \rightarrow Discriminative\ Model \rightarrow Spam/Not\ Spam A generative model can instead learn email characteristics and generate new email samples.	10	CO1	BT2

PART C: Answer Any 1 Following Questions. 10 M				
Q.N.	**Questions**	**M**	**CO**	**BT**
8	Answer Generative AI has several Data Science applications: **Data Augmentation:** Generates additional samples to increase training data. **Synthetic Data Generation:** Produces artificial datasets that resemble real data. **NLP:** Generates text, summaries or other language content. **Image Synthesis:** Creates new images based on learned visual patterns. **Recommendation Systems:** Can support data generation where interaction data is limited. Thus, Generative AI can support both data preparation and application-level tasks in Data Science.	10	CO1	BT2
9	Answer **Explicit density estimation** represents the data probability distribution directly. For example: P(X)P(X) is explicitly modeled. **Implicit density estimation** learns a procedure for generating samples without directly representing the probability density. The role of both approaches is to capture characteristics of the data distribution so that the model can represent or generate samples resembling the original data.	10	CO1	BT2

PART D: Answer Any 1 Following Questions. 10 M				
Q.N.	**Questions**	**M**	**CO**	**BT**

ANSWER KEY

10	**Answer – 10 Marks** A **Variational Autoencoder (VAE)** can be used to learn a compact representation of customer records and generate new synthetic records with similar characteristics. **1. Data Collection and Preparation – 2 Marks** First, the original customer records are collected. For example, the dataset may contain: • Age • Income • Purchase frequency • Spending amount • Customer category The data is preprocessed before giving it to the VAE. Numerical values can be normalized, and categorical values can be suitably encoded. X={x1,x2,...,xn}X = \{x_1,x_2,\ldots,x_n\} where XX represents the original customer dataset. **2. Encoder – 2 Marks** The preprocessed customer record is given to the **encoder**. The encoder converts the input record into parameters of a probability distribution in the latent space: x→Encoder→μ,σx \rightarrow Encoder \rightarrow \mu,\sigma where: • μ\mu = mean of latent distribution • σ\sigma = standard deviation of latent distribution The encoder therefore provides a compact representation of the important characteristics of the customer record. **3. Latent Representation – 2 Marks** The latent representation is obtained by sampling from the distribution described by μ\mu and σ\sigma. Using the reparameterization approach: z=μ+σ⊙ϵz=\mu+\sigma\odot\epsilon where: ϵ∼N(0,I)\epsilon \sim N(0,I) Here, zz is the **latent representation**. The latent space captures important patterns present in the customer data. **4. Decoder and Reconstruction – 2 Marks** The latent vector zz is passed to the decoder: z→Decoder→x^z \rightarrow Decoder \rightarrow \hat{x} The decoder reconstructs the customer record as x^\hat{x}. The reconstructed record is compared with the original record to calculate reconstruction error. The VAE also uses KL divergence to maintain a meaningful latent distribution. The training objective is generally represented as: LVAE=Lreconstruction+LKLL_{VAE}=L_{reconstruction}+L_{KL} **5. Generation of New Customer Records – 2 Marks** After training, the VAE can generate new customer records by selecting or sampling a new latent vector from the learned latent space. znew→Decoder→xnewz_{new} \rightarrow Decoder \rightarrow x_{new} The decoder converts this new latent vector into a synthetic customer record. **Required Workflow** Original Customer Records→Preprocessing→Encoder→μ,σ→Latent Space→Decoder→Synthetic Customer Records\boxed{ Original\ Customer\ Records \rightarrow Preprocessing \rightarrow Encoder \rightarrow \mu,\sigma \rightarrow Latent\ Space \rightarrow Decoder \rightarrow Synthetic\ Customer\ Records } **Selected Stage for Generating New Samples** The **latent space is the key stage used for generating new samples**. After training, a new latent vector zz is sampled from the learned latent distribution and supplied to the decoder. The decoder then produces a new synthetic customer record. **Conclusion** Thus, a VAE first learns the structure of existing customer records through the encoder and latent space and then uses the decoder to reconstruct or generate new records. This makes the VAE suitable for synthetic customer-data generation.	10	CO2	BT3
11	answer The complete flow is: Input x→Encoder→μ,σ→Reparameterization→z→Decoder→x^Input\ x \rightarrow Encoder \rightarrow \mu,\sigma \rightarrow Reparameterization \rightarrow z \rightarrow Decoder \rightarrow \hat{x} The encoder compresses the input into latent information. The reparameterization stage creates a latent sample. The decoder uses this latent representation to reconstruct or generate numerical data.	10	CO2	BT3

PART E: Answer Any 1 Following Questions. 10 M				
Q.N.	**Questions**	**M**	**CO**	**BT**
12	answer The required equation is: z=μ+σϵz=\mu+\sigma\epsilon where ϵ∼N(0,I)\epsilon\sim N(0,I) This transformation separates the randomness from the learnable parameters. Therefore, gradients can be propagated through μ\mu and σ\sigma, enabling gradient-based training of the VAE.	10	CO2	BT3

ANSWER KEY

13	Answer The training flow is: z→G(z)→Fake Imagez\rightarrow G(z)\rightarrow Fake\ Image and Real Image→D(x)Real\ Image\rightarrow D(x) while: Fake Image→D(G(z))Fake\ Image\rightarrow D(G(z)) **Generator GG:** Produces synthetic photographs. **Discriminator DD:** Determines whether an image is real or generated. The generator improves its ability to produce realistic images while the discriminator improves its ability to distinguish real and fake images.	10	CO2	BT3

ANSWER KEY