

Roll No .....

**IO-702 (D) (GS)**  
**B.Tech., VII Semester**  
 Examination, November 2023  
**Grading System (GS)**  
**Machine Learning**  
 Time : Three Hours

Maximum Marks : 70

Note: i) Attempt any five questions.

किन्हीं पाँच प्रश्नों को हल कीजिए।

ii) All questions carry equal marks.

सभी प्रश्नों के समान अंक हैं।

iii) In case of any doubt or dispute the English version question should be treated as final.

किसी भी प्रकार के संदेह अथवा विवाद की स्थिति में अंग्रेजी भाषा के प्रश्न को अंतिम माना जायेगा।

1. a) Differentiate between batch learning and online learning with examples. 7

उदाहरण सहित बैच लर्निंग और ऑनलाइन लर्निंग के बीच अंतर बताइए।

b) Explain about distribution of data in terms of hypothesis and hypothesis space in 2D co-ordinate plane with an example. 7

एक उदाहरण के साथ 2D समन्वय विमान में परिकल्पना और परिकल्पना स्थान के संदर्भ में डाटा के वितरण के बारे में बताइए।

2. a) Illustrate the major tasks involved in data preprocessing in machine learning. 7

मशीन लर्निंग में डाटा प्रीप्रोसेसिंग में शामिल प्रमुख कार्यों का वर्णन करें।

b) Design a dendrogram using complete linkage method for the given data points. 7

	R	S
P <sub>1</sub>	0.41	0.73
P <sub>2</sub>	0.39	0.57
P <sub>3</sub>	0.35	0.44
P <sub>4</sub>	0.33	0.22
P <sub>5</sub>	0.43	0.31
P <sub>6</sub>	0.36	0.40

दिए गए डाटा बिंदुओं के लिए पूर्ण लिंकेज विधि का उपयोग करके एक डेंड्रोग्राम डिजाइन करें।

	R	S
P <sub>1</sub>	0.41	0.73
P <sub>2</sub>	0.39	0.57
P <sub>3</sub>	0.35	0.44
P <sub>4</sub>	0.33	0.22
P <sub>5</sub>	0.43	0.31
P <sub>6</sub>	0.36	0.40

3. a) Construct a tree by using BIRCH algorithm for the following data set  $D = \{(1,2), (2,4), (3,5), (4,6), (5,6), (5,9), (6,7), (6,8), (7,2), (8,9)\}$  the branching factor  $B$  is 3 maximum number of sub cluster in each leaf node  $L$  is 5 and threshold on diameter of sub cluster stored in leaf node is 2.6 . 7

निम्नलिखित डाटा सेट के लिए BIRCH एल्गोरिथम का उपयोग करके एक पेड़ का निर्माण करें  $D = \{(1,2), (2,4), (3,5), (4,6), (5,6), (5,9), (6,7), (6,8), (7,2), (8,9)\}$  शाखा कारक  $B$  3 है प्रत्येक पत्ती नोड  $L$  में उप क्लस्टर की अधिकतम संख्या 5 है और पत्ती नोड में संग्रहीत उप क्लस्टर के व्यास पर सीमा 2.6 है।

- b) How do you represent a Gaussian mixture model? Explain in detail about expectation maximization in Gaussian mixture models. 7

आप गॉसियन मिश्रण मॉडल का प्रतिनिधित्व कैसे करते हैं? गॉसियन मिश्रण मॉडल में अपेक्षा अधिकतमीकरण के बारे में विस्तार से बताइए।

4. a) Determine the output  $y$  of a three input neuron with bias. The input feature vector is  $(p_1, p_2, p_3) = (0.6, 0.8, 0.5)$  and weight values are  $[w_1, w_2, w_3, b] = [0.3, 0.4, -0.17, 0.43]$ . Use binary sigmoid function as activation function. 7

पूर्वाग्रह के साथ तीन इनपुट न्यूरॉन का आउटपूट  $y$  निर्धारित करें। इनपुट फीचर वेक्टर  $(p_1, p_2, p_3) = (0.6, 0.8, 0.5)$  है और वजन मान  $[w_1, w_2, w_3, b] = [0.3, 0.4, -0.17, 0.43]$  हैं। सक्रियण फंक्शन के रूप में बाइनरी सिग्मॉइड फंक्शन का उपयोग करें।

- b) Consider the following dataset , apply Bernouli's Navie Bayes and predict whether a person has a disease or not based on their age, sex, and fever. 7

निम्नलिखित डाटासेट पर विचार करें बर्नौली के नेवी बेज को लागू करें और भविष्यवाणी करें कि किसी व्यक्ति को उनकी उम्र, लिंग और बुखार के आधार पर कोई बीमारी है या नहीं।

Person	Sex	Fever	Disease
Yes	Male	Yes	False
Yes	Female	No	True
No	Male	Yes	False
Yes	Female	No	True
No	Male	Yes	False
Yes	Male	No	True

5. a) Consider the following confusion matrix and calculate performance measure for accuracy, precision and recall. 7

निम्नलिखित भ्रम मैट्रिक्स पर विचार करें और सटीकता, परिशुद्धता और रिकॉल के लिए प्रदर्शन माप की गणना करें।

Category	English people	Greek people
English people	67	23
Greek people	16	75

- b) Suppose one hundred people rated the alpha release of a travel and tourism app on a scale of 1 to 10, where 25 people gave a rating of 1, 36 people gave a rating of 2, 21 people gave a rating of 3, 8 people gave a rating of 4, and 10 people gave a rating of 5. Find out the weighted average. 7  
मान लीजिए कि एक सौ लोगों ने एक यात्रा और पर्यटन ऐप की अल्फा रिलीज को 1 से 10 के पैमाने पर रेटिंग दी, जहाँ 25 लोगों ने 1 रेटिंग दी, 36 लोगों ने 2 रेटिंग दी, 21 लोगों ने 3 रेटिंग दी, 8 लोगों ने 4 की रेटिंग दी गई, और 10 लोगों ने 5 रेटिंग दी। भारित औसत ज्ञात कीजिए।
6. a) Discuss with random subspace ensembles hyper parameters with an example. 7  
एक उदाहरण के साथ यादृच्छिक उप-स्थान संयोजन हाइपर पैरामीटर्स पर चर्चा करें।
- b) Explain about the working of the Random Forest algorithm with a neat sketch. 7  
एक स्वच्छ रेखाचित्र के साथ रैंडम फॉरेस्ट एल्गोरिथम की कार्यप्रणाली के बारे में समझाइए।
7. a) Consider a PCA on a 800-dimensional dataset and the variance ratio is 98%. How many dimensions will the resulting dataset have? Explain. 7  
800-आयामी डाटासेट पर एक PCA पर विचार करें और विचरण अनुपात 98% है। परिणामी डाटासेट में कितने आयाम होंगे? व्याख्या करें।

- b) Distinguish between PAC and VC with suitable examples. 7

उपयुक्त उदाहरणों सहित PAC और VC के बीच अंतर स्पष्ट करें।

8. Write a short note on any two of the following : 14

- a) Data Science  
b) Parameters estimations - MAP  
c) Gradient Boosting  
d) Variance Ratio

निम्नलिखित में से किन्हीं दो पर एक संक्षिप्त टिप्पणी लिखें।

- अ) डाटा विज्ञान  
ब) पैरामीटर अनुमान - MAP  
स) ग्रेडिएंट बूस्टिंग  
द) विचरण अनुपात

\*\*\*\*\*