

Question Paper

Exam Date & Time: 08-Apr-2024 (02:30 PM - 04:00 PM)



MANIPAL ACADEMY OF HIGHER EDUCATION

Machine Learning Principles and Applications [AML 5203]

Marks: 50

Duration: 90 mins.

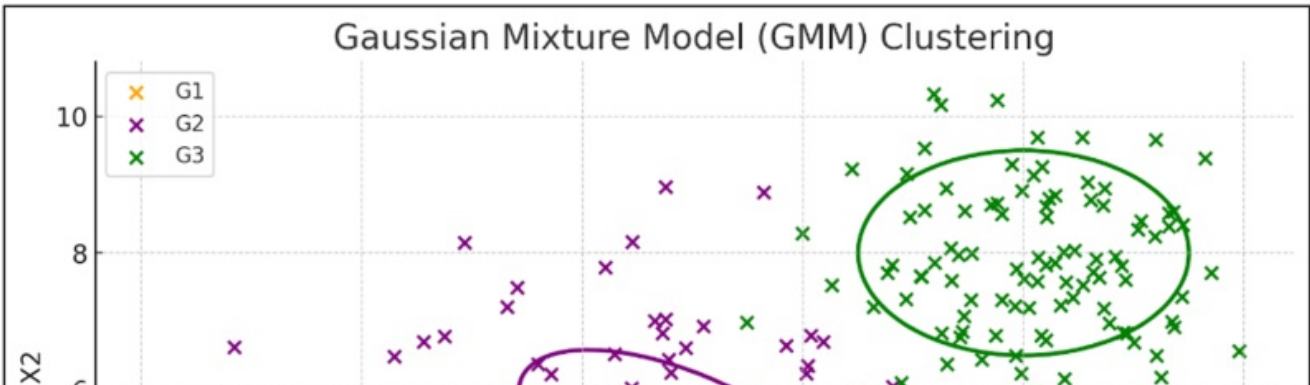
Keep answers short and precise

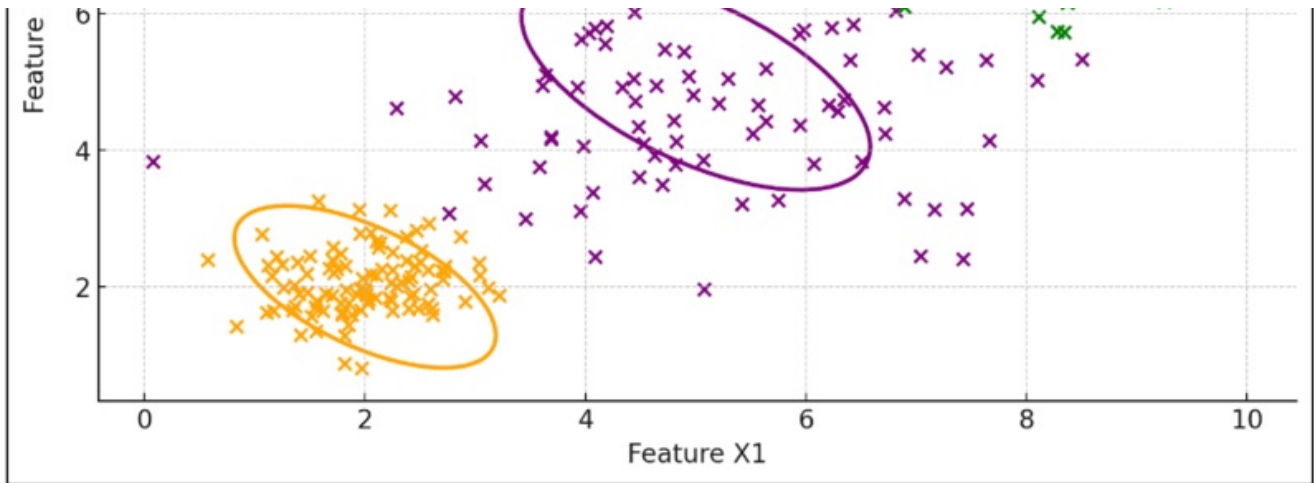
Answer all the questions.

- 1) [CO1, L5] Say true or false and justify your response: (10)
1. The conditional independence assumption of Naive Bayes makes it inefficient for real-world datasets where features are often dependent..
 2. The Gaussian Naive Bayes classifier is best suited for continuous features.
 3. Feature scaling is not necessary before applying Naive Bayes classifier due to its probabilistic nature.
 4. Naive Bayes classifiers can be used for regression tasks by predicting numerical values instead of class labels.
 5. Laplace smoothing is only applicable to Naive Bayes classifiers used in text classification tasks.
- 2) [CO2, L3] Given the following dataset, predict whether a new team member with **High** time spent on project, **Average** test scores, and **Good** participation will lead to a successful project using a Naive Bayes classifier: (10)

Time Spent on Project	Test Scores	Participation	Project Success
Low	High	Average	No
High	Low	Good	Yes
Average	<input type="text" value="Average"/>	None	No
High	Average	Excellent	Yes
High	Low	Good	No

- 3) [CO3, L5] The following diagram represents a dataset plotted in a two-dimensional space, with points that seem to cluster into three distinct groups. A Gaussian Mixture Model (GMM) has been fitted to this data, resulting in three Gaussian distributions, each representing a cluster: (10)





1. Which Gaussian (G1, G2, or G3) appears to have the highest variance? Explain how you can tell based on the diagram.
2. If a new data point is added to the plot and falls within the overlap between G1 and G2, how does the GMM decide which cluster to assign the point to? Discuss the concept of "soft clustering" in your explanation.

4) [CO3, L5] Say true or false and justify your response: (10)

1. The "spherical" covariance type in a GMM assumes all features vary independently but with the same variance across each dimension.
2. GMMs are superior to K-means clustering when dealing with elongated or irregularly shaped clusters.
3. Using GMM for classification tasks is common practice because it naturally separates data into distinct groups.
4. A Gaussian Mixture Model (GMM) is capable of modeling data that comes from multiple Gaussian distributions within the same dataset.
5. GMMs perform well with high-dimensional data without any modifications.

5) [CO2, L4] (10)

1. Explain the efficiency of Gaussian mixture models in dealing with outliers.
2. How can we use Gaussian mixture models for generating data?

-----End-----