Parul University

Faculty of Engineering and Technology

Parul Institute of Technology

Machine Learning (303105353)

Assignment No. - 02

---

1. A data scientist is developing a system to predict car prices based on features like **age, mileage, engine size, and brand reputation score**. When plotting the data, she notices that the relationship between price and mileage is **curved rather than straight**.

   Question:

   Explain whether a **linear** or **non-linear** model would be more appropriate for this prediction task. What advantages and trade-offs does each model type offer in this scenario?

2. A photo management app automatically tags uploaded pictures. A single image might belong to **multiple categories** (e.g., "Person," "Outdoor," "Night"), while in other cases, each image belongs to **only one** (e.g., "Cat," "Dog," "Bird").

   Question:

   Differentiate between **multi-class** and **multi-label classification** in this context. Suggest one algorithm suitable for each and justify your choices.

3. A real estate company collects the following data to predict house prices:

Size (sq.ft)	Bedrooms	Distance to City (km)	Price (₹ Lakhs)
1000	2	5	40
1500	3	3	65
2000	3	8	70

   Question:
   (a) Explain how a **simple linear regression** model (using only Size) would estimate prices.
   (b) Describe how a **multilinear regression** model could improve predictions by using all three features.
   (c) Discuss one practical limitation of linear regression for this dataset.

4. A bank wants to predict whether a loan applicant will **default** or **repay** based on three attributes: **Income (High/Low)**, **Credit History (Good/Bad)**, and **Employment Status (Stable/Unstable)**. The training data shows that most defaulters have low income and unstable jobs.

   Question:

Explain how the **Naïve Bayes Classifier** would make predictions for a new applicant with **Low Income**, **Good Credit History**, and **Unstable Employment**. Why does Naïve Bayes assume independence among attributes, and what are the implications of this assumption?

5. A telecom company uses a **Decision Tree model** to predict customer churn (Yes/No) based on features like **Contract Type**, **Monthly Charges**, and **Customer Support Calls**. After testing, the model achieves **92% accuracy on training data** but only **78% on test data**.

   Question:

   (a) Explain how the **ID3** or **CART** algorithm would construct such a tree (mention splitting criteria like Information Gain or Gini Index).
   (b) Discuss what this drop in accuracy suggests about **error bounds** and model generalization.
   (c) Suggest one technique to reduce overfitting in this decision tree.

Date:25/12/2025

Dr. Vinod Patidar
Asso. Prof. (CSE Dept)
PIT, Vadodara

Note: Submit the assignment on or before 01/01/2026.