ourdiu.com | Shared by Abdullah Tawsif (CSE-232)

Machine Learning

Q6. A retail company wants to predict customer monthly spending based on age and annual income.

Cufstomer	Age	Annual Income (K USD)	Monthly Spending (USD)
1	22	25	300
2	28	35	450
3	35	50	700
4	40	60	850

Using the given historical data, estimate the monthly spending of a 45-year-old customer whose annual income is 70K USD. Show the necessary calculations and briefly interpret the predicted value in the context of the business problem.

Q7. An ICU patient monitoring system continuously logs three vital indicators to estimate critical illness risk:

- **Patient Age (X1):** Spans 20 to 85 years (uniform distribution).
- **Heart Rate (X2):** Spans 50 to 180 bpm (normal distribution).
- **Blood Biomarker (X3):** Spans 0.001 to 2500.0 mg/dL (heavily right-skewed with severe clinical spikes).

The system evaluates model error using the Mean Squared Error (MSE) cost function J(w,b) = (yi - (w * xi + b))^2. During model fitting with Gradient Descent, the engineer observes that the loss rapidly shoots up toward infinity.

Develop a feature preprocessing pipeline for variable X3 that enables stable optimization alongside X1 and X2. Justify your pipeline components.

Quiz 1
Sec D
Fall 26
Shahriar Shakil sir

ourdiu.com | Shared by Abdullah Tawsif (CSE-232)

Machine Learning

6+4

Q6. A fintech firm is developing two models:

- **Property Valuation Model:** Predicts home price Y (in $100,000) using living space X1 (thousands of sq. ft.) and X2 (hundreds of sq. m.). Historical sample data:

Property	X1 (k sq ft)	X2 (h sq m)	Y (Price $100k)
House A	1	2	5
House B	2	3	7
House C	3	4	9

- **Credit Risk Model:** Assesses loan applicants. Data inspection shows missing values in two fields:
  - **Annual Income:** Missing mainly for applicants with very high earnings, because some choose not to disclose it.
  - **Home Ownership:** Missing in about 5% of records because of occasional network timeouts during online application submission.

A. Demonstrate mathematically why standard multivariable linear regression fails on the property valuation dataset. Identify the underlying issue and propose two production solutions.

B. Analyze the missing data patterns in the credit risk dataset and evaluate the business impact of applying simple mean imputation to the missing income values

Quiz 1
Sec I
Fall 26
Shahriar Shakil sir