Unsupervised Learning Lab

Assignment 2

PCA – Dimensionality Reduction algorithm

Synopsis of Dataset

What is this dataset?

This dataset contains information about **8,950 credit card customers** with 18 different features describing their spending behaviour and account details over a 6–12-month period.

What the columns mean in simple terms:

Account Information

- **CUST_ID**: A unique ID for each customer

- **TENURE**: How long the customer has had the account (6-12 months)

Money the customer owes/spends

- **BALANCE**: How much money they owe on their credit card

- **CREDIT_LIMIT**: Their maximum spending limit

- **PURCHASES**: Total amount spent on purchases

- **ONEOFF_PURCHASES**: Purchases paid in one payment

- **INSTALLMENTS_PURCHASES**: Purchases paid in instalments

- **CASH_ADVANCE**: Cash taken from credit card (like a loan)

How often they do things

- **PURCHASES_FREQUENCY**: How frequently they make purchases

- **ONEOFF_PURCHASES_FREQUENCY**: How often they make one-time payments

- **PURCHASES_INSTALLMENTS_FREQUENCY**: How often they use instalment plans

- **CASH_ADVANCE_FREQUENCY**: How often they take cash advances

- **CASH_ADVANCE_TRX**: Number of cash advance transactions

- **PURCHASES_TRX**: Number of purchase transactions

Payment behaviour

- **PAYMENTS**: Amount they've paid back

- **MINIMUM_PAYMENTS**: The minimum amount they're required to pay

- **PRC_FULL_PAYMENT**: How often they pay the full balance (0 = never, 1 = always)

Key observations from the data:

- Most customers have had their cards for about 12 months

- There's a huge range in spending - from $0 to nearly $50,000

- Some customers never use certain features (like cash advances)

- About 300 customers are missing minimum payment information

- The average credit limit is around $4,500, but goes up to $30,000