### **Explanation of the Assignment in Simple Terms:**

### **Dataset Overview**

* You have a dataset called advertisement.csv that includes data on:
  + TV advertisement budgets,
  + Radio advertisement budgets,
  + Newspaper advertisement budgets,
  + and Sales figures (dependent variable).

The goal is to **predict Sales** based on advertisement spending.

### **Explanation of the Code**

#### **1. Initialization**

* The workspace is cleared using clear, close all, and clc commands.
* The data is loaded using readmatrix from an Excel file (advertisement.xlsx), which contains:
  + TV: Advertisement spending on TV.
  + Radio: Advertisement spending on Radio.
  + Newspaper: Advertisement spending on Newspapers.
  + Sales: Resulting sales.
* Individual columns are extracted for easy manipulation and analysis.

#### **2. Data Visualization**

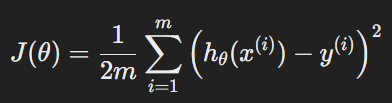
* Scatter plots are created to visualize relationships between:
  + TV and Sales.
  + Radio and Sales.
  + Newspaper and Sales.
* Each subplot provides an intuitive view of how advertising spending in each medium affects sales.

#### **3. Feature Normalization**

* **Why Normalize?** Normalization is essential to scale the features to a similar range, ensuring that gradient descent converges efficiently.
* **Normalization Steps:**
  + Subtract the mean (mu) of each feature and divide by its standard deviation (sigma).
  + Sales data is also normalized for consistency.
* The normalized data is used for gradient descent.

#### **4. Cost Function**

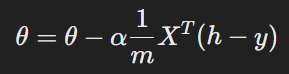
* A custom cost function ComputeCost is implemented to calculate the Mean Squared Error (MSE):



* + h = X \* theta: Hypothesis or predicted values.
  + m: Number of training examples.
  + a: Difference between predictions and actual values.

#### **5. Gradient Descent**

* A custom gradient descent function GradientDescent is implemented to minimize the cost function iteratively:



* + alpha: Learning rate (set to 0.01 for this example).
  + iterations: Number of iterations for convergence (set to 500).
* The initial cost is computed, and parameters (theta) are updated in each iteration to minimize the cost.

#### **6. Correlation Analysis**

* **Why Check Correlation?** High correlation between features (multicollinearity) can affect the model's interpretability and accuracy.
* The **correlation matrix** is calculated to measure the relationships between:
  + TV and Radio.
  + Radio and Newspaper.
  + Newspaper and TV.
* If the correlation between any two features exceeds the threshold (0.8 in this case), one of the features is dropped. For example:
  + If Radio and Newspaper have a high correlation, Newspaper is removed from the feature set.

#### **7. Model Training**

* After dropping correlated features (if necessary), the model is retrained:
  + A new feature matrix X is created by adding a bias term (column of ones) to the features.
  + The features are re-normalized.
  + Gradient descent is run again to update theta values.

#### **8. Denormalizing Parameters**

* To interpret the model in the original scale, the coefficients (theta) are denormalized:
  + The intercept is adjusted using the mean and standard deviation of y.
  + The feature coefficients are scaled back to their original units.

#### **9. Final Output**

* The final model is presented in the form:



* theta\_0: Intercept.
* theta\_1, theta\_2: Coefficients for TV and Radio, respectively.

#### **Questions and Solutions**

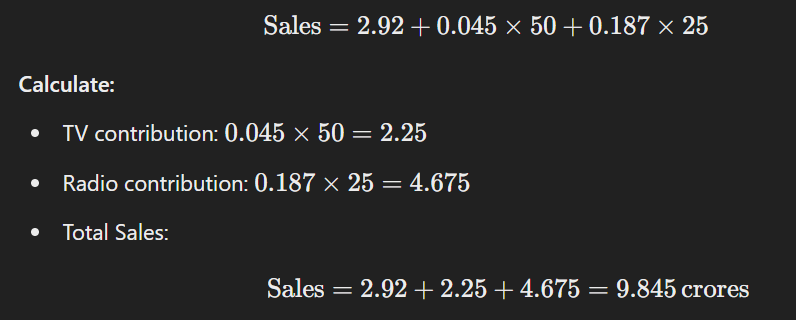
##### **Q1: What happens if TV marketing increases by 2 lakhs?**

* The coefficient of TV is **0.045**. This means **1 lakh increase in TV budget increases Sales by 0.045 crores**.
* If TV increases by **2 lakhs**:



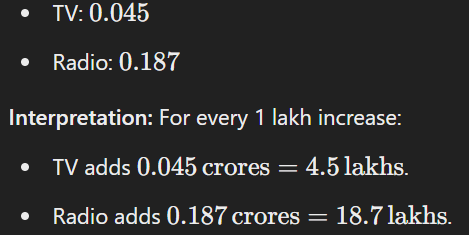
##### **Q2: What are the Sales when TV is 50 lakhs, and Radio is 25 lakhs?**

* Substitute the given values into the model:



##### **Q3: Which Advertisement Channel Gives the Best ROI?**

* **Return on Investment (ROI):** Compare the coefficients:



**Conclusion:** Radio is more effective for increasing Sales per unit budget and offers better ROI.