Machine Learning Laboratory
Assignment 1

Department of Computer Science and Engineering
CS4092D Machine Learning Laboratory

This dataset contains data of customers who buy clothes online. The store offers in-store style and clothing advice sessions. Customers come into the store, have sessions/meetings with a personal stylist, and then they can go home and order either on a mobile app or website for the clothes they want. The company is trying to decide whether to focus their efforts on their mobile app experience or their website.

1. Identify and select the 1 or 2 most important features from the dataset that are likely to influence the *Yearly Amount Spent*. Justify your selection with reasoning or through exploratory data analysis, such as a correlation matrix.

2. Model the output variable *Yearly Amount Spent* as a linear combination of the selected features. Write down the hypothesis equation.

3. Formulate the cost function for the linear regression model.

4. Write a function to implement gradient descent that optimizes the cost function. Use this function to find the optimal values of the weight vector (model parameters). Experiment with different learning rates and report the one that provides the best results.

5. Output the optimum values of the weight vector, the intercept term, and the chosen hyperparameters, such as the learning rate and number of iterations.

6. Plot the regression line (or plane if using 2 features) on the training data. If two features are used, create a 3D plot showing the plane.

7. Plot the training curve that shows the cost function value versus the number of iterations (epochs).

8. Generate 2 sample test cases (new data points) and predict the *Yearly Amount Spent* for these cases using the trained model.