Data Science CSF20

Assignment 1: Training a Linear Regression Model

Task: In this assignment, you are going to train a linear regression model on the given data set. **(100 points)**

About Dataset:

You are given two CSV files

I. Train.csv
II. Test.csv

Both files contain two columns (x as the input variable, and y as the output variable). You are required to train your model using the ***train.csv*** file and then test your model on the ***test.csv*** file.

You can access the dataset from the given link: https://drive.google.com/drive/folders/1PH_C-RnXba9HvX66QfScPU2VQJTapFop?usp=sharing

Instructions:

1. Create a Google Collab Notebook and name it as A1-Roll_Number (A1- BITF20M002) [1]
2. Connect your collab Notebook file with Google Drive.[1]
3. Import necessary libraries such as numpy, pandas, etc. [2]
4. Load the datasets and display 10 records from both files using the head() function. (Use the pandas' library for this) [6]
5. Convert pandas' data-frames (obtained in step 4) to numpy arrays and display their shapes. [10]
6. **Training: [45]**
   Your goal is to minimize the cost by finding the optimal values of $theta0$ and $theta1$ within a given range (-0.5 to 1.5) using a nested loop. The values for $theta0$ and $theta1$ should be incremented by 0.2 at each iteration, starting from -0.5. The hypothesis function ($y' = h\theta(x) = \theta_0 + \theta_1$) should be implemented.
   Compute the cost using:

   $$J(\vartheta_0,\vartheta_1) = \frac{1}{2m}\sum_{i=1}^{m}(y'^{(i)} - y^{(i)})^2$$

   And achieve the objective function by:

   $$\min_{\vartheta_0,\vartheta_1} J(\vartheta_0,\vartheta_1)$$

7. **Testing:** Predict the output for the test dataset using the parameters calculated in step 6. [25]
8. Draw Graphs of the predicted output and actual output.[10]

Bonus Task: [50]

Randomly initialize values for $theta0$ and $theta1$, and apply the gradient descent algorithm to iteratively update the values and find the optimal values where the cost function is minimized.

Sample Collab Notebook:

https://colab.research.google.com/drive/1pagcPnxOy2HU2ACFAlSl8h84IiBkXfy2?usp=sharing

Download and then submit your A1-Roll_Number.ipynb file