DATA SCIENCE ASSIGNMENT

Introduction

- Lithium-ion batteries are becoming ubiquitous in our lives, from consumer electronics to EVs. In this assignment, you will work on time series data of these batteries in two different situations. This assignment consists of two parts - Part 1 & Part 2.
- Part 1 - Feature extraction from battery data
    - You are provided with the data collected from a battery that has been operational in an electric vehicle.
    - It is a time-series data sampled at a variable rate.
    - The data is provided as a csv file.
    - You are required to complete three tasks in this part.
- Part 2 - Modelling the SOH(State Of Health) of a cell
    - Here, the dataset provided shows how the capacity of a cell varies with cycle count.
    - You are required to model the SOH of a cell from the data collected in the laboratory.
    - The data is provided as a csv file.
    - You are required to complete three tasks in this part.
- We have provided the definitions of the basic concepts. Please feel free to use Google to understand them further.
- Submission :
    - Use Python and Google Colaboratory to write your code.
    - Submit your response as a link to your colab file.
    - Complete each task in a separate block and add comments to your code wherever necessary.
    - Make use of the standard packages for data processing and regression. Clearly document any external packages used by your code.
    - At the end of the notebook, include a section as a short write-up with subsections for each part and sub-subsections for each task, to present your answers and explanations or anything else of interest.

DATA SCIENCE ASSIGNMENT

Part 1 - Feature extraction

Definitions:

1. SOC (State Of Charge) : It is the level of charge in a battery relative to its capacity. A battery is fully charged when it is at 100% SOC and fully discharged when it is at 0% SOC.

2. FEC (Full Equivalent Cycles) : FEC can be defined as the number of **full** charge-discharge cycles that a battery has undergone. When a battery gets charged from 20% SOC to 70% SOC and then discharged back to 20% SOC, its FEC does not increase by 1. Instead, it takes two such “half” cycles for the FEC to go up by 1. This is because the FEC is measured based on the charge throughput of the battery; it increases by unity only when the entire capacity of the battery gets drained from it and then pumped back in again regardless of the number of steps. (Read this short description for more clarity - refer to section 2.2)

Description of the dataset:

1. There are 4 features : timestamp, soc, voltage, current.
2. The units of voltage and current are Volts and Amperes respectively.

Tasks:

1. Perform data preprocessing - remove null values, look for outliers and clean up the data.
2. Based on SOC and current, create an algorithm to **obtain the FECs that the given battery has undergone**. Present the obtained FEC along with your reasoning in the write up section. Please note that the 'current' column does not discriminate between charging current and discharging current. (hint : add a new feature – ‘charge current’ by monitoring the change in SOC).
3. The capacity of a battery is measured in Ampere hours (Ah). Is it possible to obtain the capacity of the battery from the given information? If yes, calculate the capacity of the battery as an average over the first ten cycles.

DATA SCIENCE ASSIGNMENT

Part 2 - Regression

Definitions:

1. RUL (Remaining Useful Life) : The capacity of a battery fades over time and cycles due to various degradation mechanisms that occur within it. Typically, a battery is retired from use when its capacity drops below 80% of its original capacity. **At a given point during its lifetime, the remaining number of cycles that can be extracted from the battery can be regarded as the RUL of the battery**.

2. SOH (State Of Health) : It is the ratio of a battery's present capacity to its initial capacity.

   SOH = current capacity/initial capacity.

Description of the dataset:

1. The dataset consists of two columns : Cycle_Index, Discharge_Capacity (Ah)
2. Cycle_Index represents the FEC count of the battery and Discharge_Capacity (Ah) is the capacity measured while discharging the battery from 100% SOC to 0%.

Tasks:

1. Model the SOH of the cell as a function of cycle number
   a. Explain your model clearly in the write-up section.
   b. Describe how your code ran on the dataset: parameters used, time taken etc.
   c. Present your model as a function named 'Predict_SOH' that takes a pandas dataframe with the same column names as the provided dataset as the argument and returns SOH values as an array. Define the function as follows:

```
def Predict_SOH(dataframe)
    .....
    return [SOH]
```

2. Using the model, determine the RUL of the battery at 500, 1000, 2000, 3000 and 3500 cycles. Plot the predicted values and the actual values and add this plot with proper title and legend to the write-up section. Assume that 80% SOH indicates that the battery has reached its end of life.
3. Write a function to evaluate the performance of the model, named 'Estimate_error' which takes a pandas dataframe with the same column names as the provided dataset as the argument and returns the tuple (RMSE, MAE), where RMSE is Root Mean Squared Error and MAE is Mean Absolute Error of the model. Define the function as follows:

```
def Estimate_error(dataframe)
    .....
    return (RMSE, MAE)
```