**[MLDL for DS] Programming Assignment 03**

**Duedate: May 20th at midnight (23:59:59)**

1. **Data description**

The dataset is downloaded from Nasa repository [3], which file format is Matlab. Therefore, I wrote a function to convert them into pandas format. Here is what it looks like after converting.

* **Battery capacity dataframe** contains only cycle, ambient temperature, datetime, and capacity. As can be seen, the capacity of a battery (full-charge capacity) reduces after each cycle, until it reaches the end-of-life threshold, which is 70% of its initial capacity. The capacity at cycle 1 presents this value. Here are some samples:

seq,cycle,ambient\_temperature,datetime ,capacity

0 1 24 2008-04-02 15:25:41 **1.856487 => initial capacity**

1 2 24 2008-04-02 19:43:48 1.846327

2 3 24 2008-04-03 00:01:06 1.835349

3 4 24 2008-04-03 04:16:37 1.835263

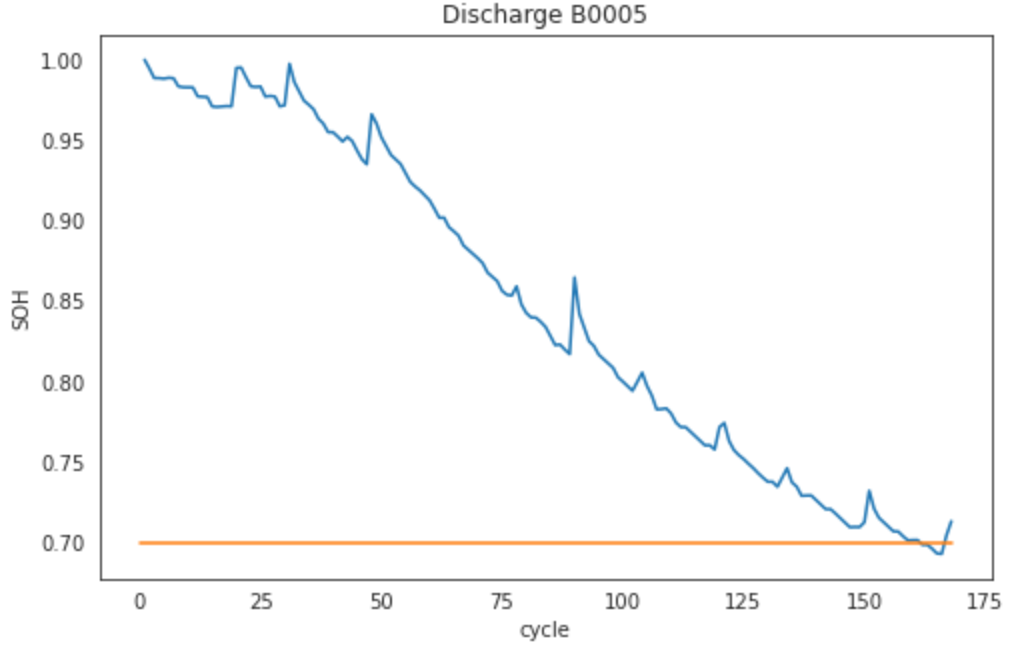
4 5 24 2008-04-03 08:33:25 1.834646

* **Usage dataframe** contains all records of battery usage throughout the time until the battery is dead. It contains the following columns: cycle, ambient\_temperature, datetime, capacity, voltage\_measured, current\_measured, temperature\_measured, current\_load, voltage\_load, time. It may require some domain-specific knowledge to understand all features clearly. Please check out papers [1,2] to find more useful information.

1. **Problems (Only need to select one out of the two)**

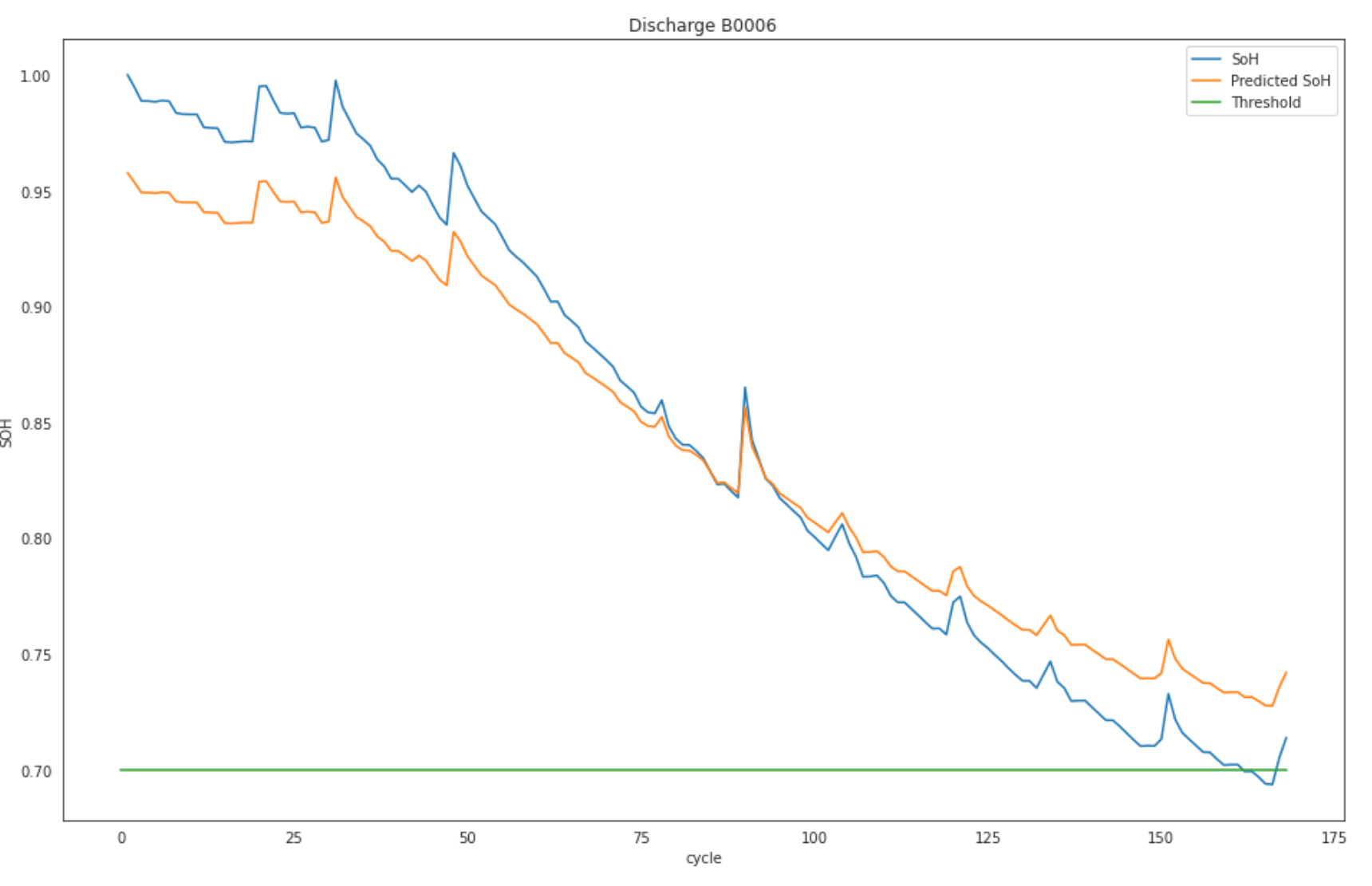
**Option 1 (100 points): Battery state-of-health (SoH) estimation**

State-of-health value is equivalent to the remaining capacity calculated by percentage. As in [2], the battery reaches the end-of-life threshold when the capacity remains only 70%. Therefore, we can estimate the SoH level based on available information of battery usage in the second dataframe.

****

We provide data frames as discussed in 1 and pre-process all necessary steps for data cleansing, data loading, and visualization. Your role is only to develop a prediction model that can estimate the SoH given historical battery usage up to the current time. We don’t limit any model selection. Therefore, all algorithms are accepted such as AdaBoost, decision tree, random forest, ANN, ensemble techniques, and so on. Besides, you can also modify data preparation functions (e.g., feature selection) if it is necessary for your model.

* Develop a model for SoH prediction: **50 points**
  + Briefly explain your model (How does your model work? Why do you select this algorithm? Which features are used and why are they used?): **30 points**
  + Your model can run and provide reasonable results: **20 points**
* Ablation study: **50 points**
  + Try three different settings with your models such as different learning rate, hidden size, number of layers and so on. **(30 points)**
  + Report and analyze results. **(20 points)**
    - What are experimental results? **(10 points)**
    - Can you explain the results? **(4 points)**
    - What are the effects of learning rate, optimizer, and the model size (size of hidden layers, number of hidden layers, so on) on results? E.g., more layers may cause overfitting, whereas less layers can lead to underfitting. (**6 points**)

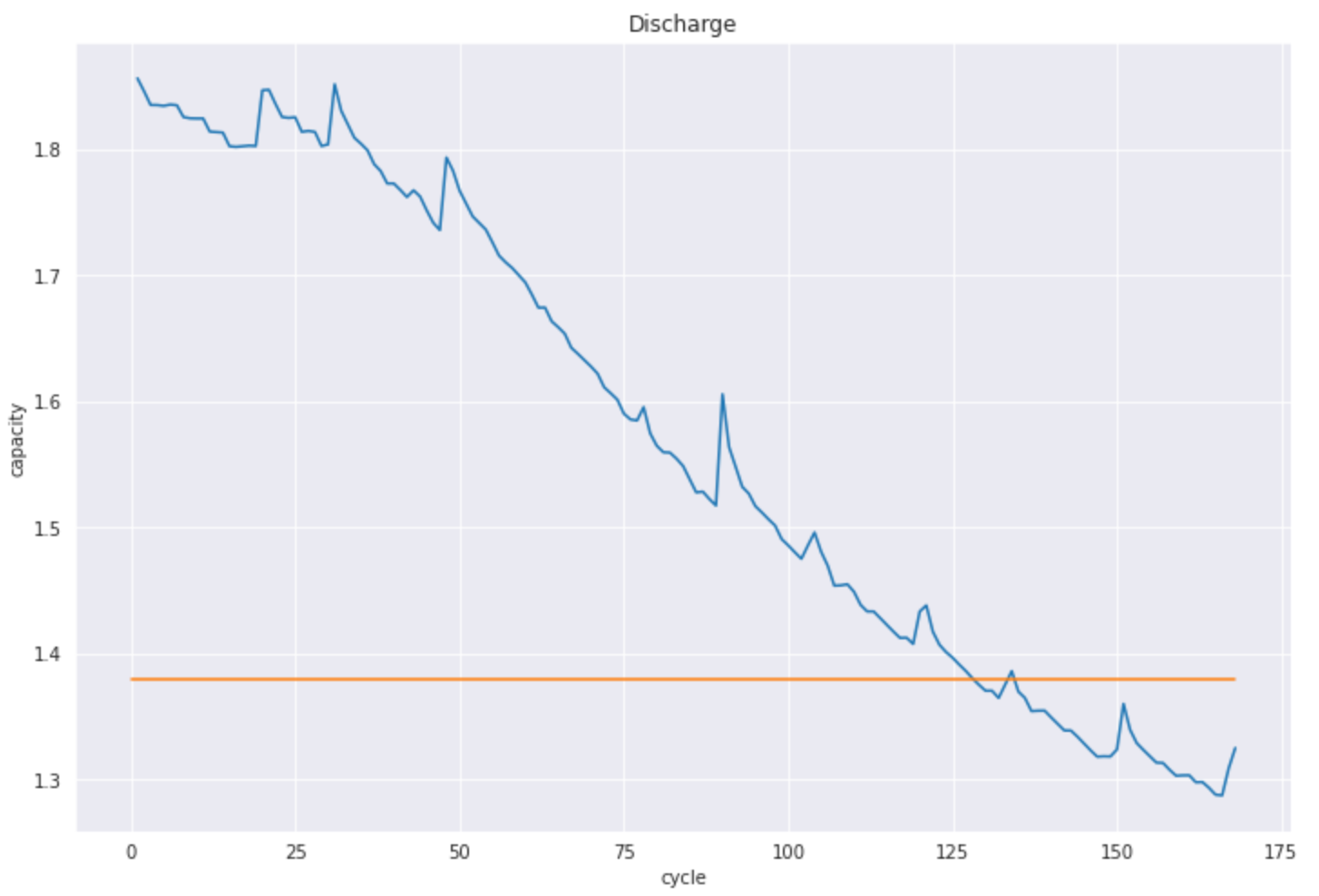
****

SoH prediction using 3 hidden layers and 1 prediction layer.

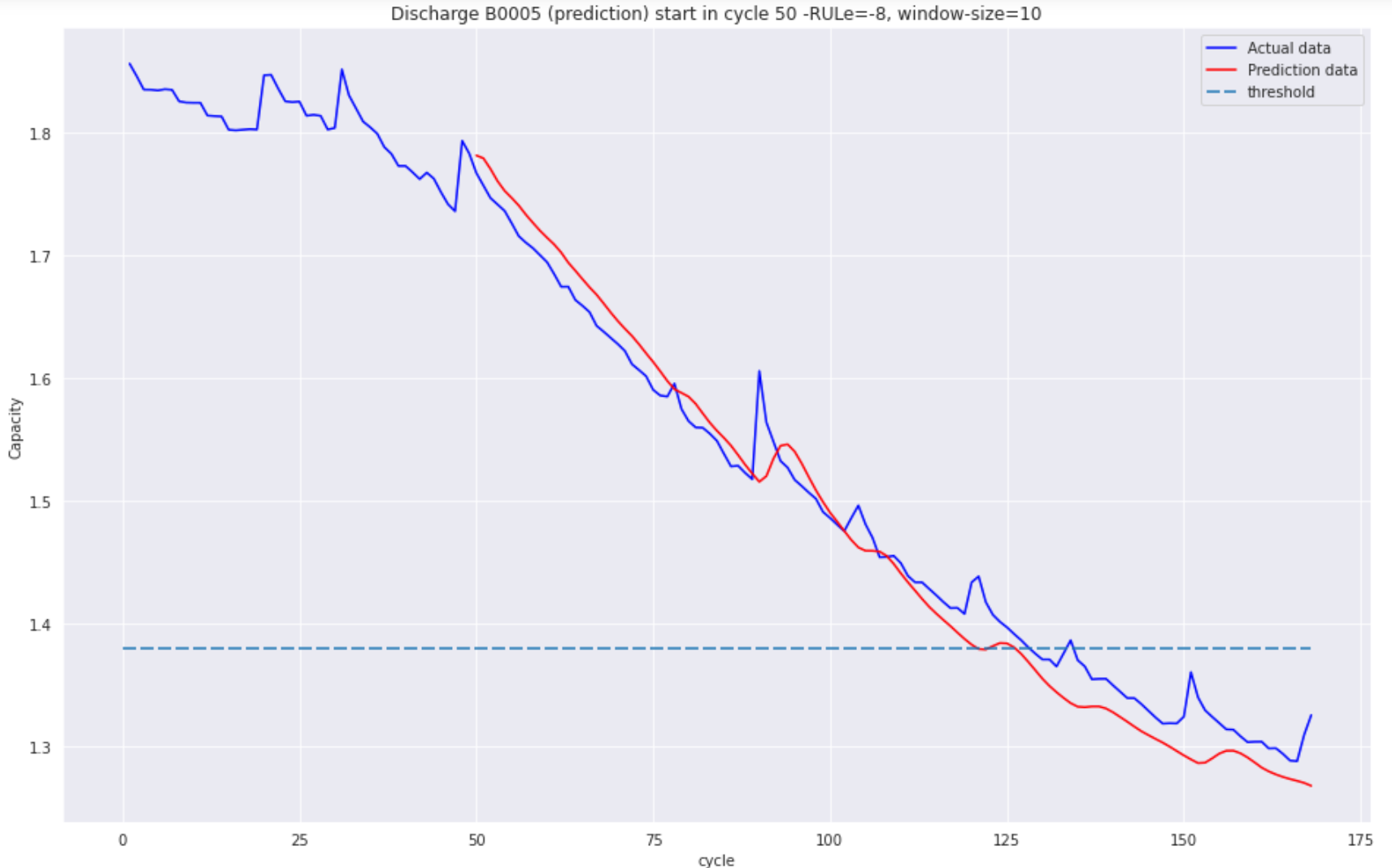
**Option 2 (100 points): Battery remaining-useful-life (RUL) estimation**

The Remaining Useful Life (RUL) is a subjective estimate of the number of remaining years that an item, component, or system is estimated to be able to function in accordance with its intended purpose before warranting replacement.

As mentioned in problem 1, the end-of-life threshold is 70% of the initial capacity that is approximately 1.38Ah. Therefore, in this problem we predict capacity values directly instead of estimating the remaining percentage. If the estimated remaining capacity of a battery is below 1.38, which means the battery is dead.

****

* Develop a model for **RUL** prediction: please refer to option 1.
* Ablation study: please refer to option 1.

****

RUL estimation using LSTM

Reference:

[1] P. Khumprom and N. Yodo, “A data-driven predictive prognostic model for lithium-ion batteries based on a deep learning algorithm,” Energies, vol. 12, no. 4, 2019.

[2] C. Wang, N. Lu, S. Wang, Y. Cheng, and B. Jiang, “Dynamic long short-term memory neural-network- based indirect remaining-useful-life prognosis for satellite Lithium-ion battery,” Appl. Sci., vol. 8, no. 11, 2018.

[3] <https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/>

[4] https://www.partneresi.com/resources/glossary/remaining-useful-life-rul

**P/S:**

* **Discussion on algorithm and program logic is welcome but sharing code with classmates is strictly prohibited. Program code must be done by yourself.**
* **You will get full points if complete implementation and code can run. Depending on the level of completion, points of each problem can be varied. Let’s submit to ETL no later than May 20th at midnight (23:59:59).**
* **For late submission, the penalty will follow the SNU’s guideline (10% deduction each day, 0 after 5 days).**
* **Should you have any questions do not hesitate to contact TA (Alex):** 
  + **Email:** [**cuongbt91@snu.ac.kr**](mailto:cuongbt91@snu.ac.kr) **Kakao Id: alexbui**