Lithium-ion Batteries Performance Prediction using Machine Learning

Dickinson DATA300 Final Project: Statistical and Machine Learning

December 18, 2024

HyunYoung Cho

Department of Data Analytics and Chemistry
Dickinson College
chohy@dickinson.edu

Olivia Petronio

Department of Data Analytics
Dickinson College
petronio@dickinson.edu

Minh Trinh

Department of Data Analytics and Mathematics
Dickinson College
trinhm@dickinson.edu

Abstract

Lithium-ion (Li-ion) batteries are prevalent today in society with their usage in rechargeable batteries such as portable electronics, electric vehicles, and grid storage. Many of the companies including Tesla (Tesla Model S), Nissan (Nissan Leaf), and NASA (hybrid-electric Boeing 787) are using Li-ion batteries for their product [1]. However, Li-ion batteries have a finite life span. Under higher temperatures, higher charge rates, or other conditions, the life span of lithium-ion batteries shortens. We used L2 regularization (Ridge) method to demonstrate the polynomial regression of lithium-ion batteries' performance, based on the data provided by Catenaro and Onori [2].

1. Introduction and Background

As discussed earlier, one of the issues of lithium-ion batteries is the adverse effect after increased temperatures. The increase in heat, increases the battery's temperature leading to the destruction of the battery, fires, and explosions. This ultimately leads people to be skeptical about its use in these applications due to their concerns about efficiency, raw material usage, and safety. One way to test lithium-ion battery performance is by measuring the amount of charge stored and delivered at the start of charge and discharge. How do different types of lithium-ion batteries store and deliver charge with varying temperature degrees? This project aims to determine which lithium-ion batteries provide the best voltage at higher temperatures for safety and efficiency. This project uses the data from Catenaro and Onori's work on "Lithium-ion batteries under galvanostatic discharge tests at different rates and temperatures".

Some technical terms regarding lithium-ion batteries are charging cutoff value, charging voltage, charging current, C-rate, and galvanostatic discharge. The charging cutoff value is the minimum voltage the battery can discharge during charging. If the voltage goes lower, it can damage the battery [3]. The charging voltage is the amount applied during the charging process [3]. The charging current is the current delivered to the battery during charging. It decreases as the battery reaches its charging cutoff value [3]. C-rate measures the rate at which a battery is completely charged or discharged, relative to its normal capacity. If it is a C-rate of 1C, it indicates the battery will need one hour to fully charge from being fully discharged [4]. Galvanostatic discharge involves the application of constant positive and negative currents to charge and discharge a system within a potential limit [5].

Polynomial regression by using L2 regularization is a powerful technique to predict *y*-variable when there are a lot of features in the dataset. Also, regularization is a technique used to prevent over-fitting in machine learning models, by reducing model complexity without excluding features. L2 regularization is based on the following equation:

$$LOSS = MSE + \lambda \sum_{j=1}^{n} \theta^2{}_j$$

Here, θ_j corresponds to the *j*-th parameter in the dataset. λ corresponds to a regularization parameter. L2 regularization adds the term $\lambda \sum_{j=1}^{n} \theta^2{}_j$ to the loss function. Unlike L1 regularization, L2 regression shrinks coefficients without forcing them to zero. The cost function or L2 regularized mean squared error is the following:

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} \left(y^{(i)} - \vec{\theta}^{\,T} \vec{x}^{(i)}\right)^2 + \frac{\lambda}{2m} \|\theta\|_2^2$$

Where:

$$\sum_{j=1}^{d} \theta_j^2 = \|\theta\|_2^2$$

This finally results in the following equation, which demonstrates the updated *j*-th parameter, θ_j'. η corresponds to learning rate, which controls how much the model parameters θ are adjusted at each step. A smaller μ means smaller updates, which can lead to slower but more stable data convergence. *m* is the number of training examples in the dataset. The $\vec{\theta}^T \vec{x}^{(i)} - y^{(i)}$ represents the error for the *i*-th training examples, which measures the difference between the predicted value and the actual value.

$$\theta_j' = \theta_j - \frac{\eta}{m} \sum_{i=1}^{m} \left(\vec{\theta}^T \vec{x}^{(i)} - y^{(i)}\right) x_j^{(i)} - \frac{\eta \lambda \theta_j}{m}$$

This finally describes how the θ_j parameter is updated at each iteration of gradient descent, considering both the error between predicted and actual outputs and L2 regularization term, which prevents overfitting of the data.

2. Experiment

The experiment was conducted with three different Li-ion battery types, which were lithium-nickel-cobalt aluminum-oxide (NCA), lithium-nickel-manganese-cobalt-oxide (NMC) and lithium-iron-phosphate (LFP). The actual battery cells used for experiment were from:

Panasonic NCR-18650B (3350 mAh), LG Chem INR21700-M50 (4850 mAh) and A123 Systems ANR26650m1-B (2500 mAh), respectively.

The experiment started from fully charged from 100% state-of-charge until the cell cutoff discharge voltage. The discharge was conducted under controlled temperature conditions, which were 5℃, 25℃, and 35℃, and varying the battery cells to galvanostatic discharge rates – C/20, 1C, 2C, 3C, and 5C for NCA, NMC, and LFP chemistry. Voltage, current, and cell surface temperature were measured on each battery cell and recorded on the data set.

For setting the temperature, the battery cell is left to reach the reference temperature as set in the chamber, so that the cell surface temperature data can be collected, starting from the initial reference temperature. Any increase or decrease in cell surface temperature can be started from the reference temperature (5℃, 25℃, and 35℃).

Figure 1. Three types of Lithium battery (NCA, NMC, and LFP battery) used in the work (*Catenaro., & Onori, 2021)*

3. Methods

Enhanced ragone plot (ERp) is useful way to visualize the data set. It is useful for evaluating the key characteristics of Li-ion batteries of how their cathode composition and operating conditions, and it can be a tool to guide energy storage system selection for applications such as electric vehicles and stationary grid storage [2]. This indicates that this plot can be a fundamental tool to determine if certain applications can function with high efficiency of Li-ion battery. The equation for specific energy for battery type *b*, cell sample *k*, and applied C-rate *x* is the following:

$$E^{b,k}(x, T_{amb}) = \frac{E^{b,k}(x, T_{amb})}{M^{b}_{nom,cell}}$$

$$E^{b,k} = \frac{C * V}{cell\ mass}$$

Here, $E^{b,k}$ is energy released by Li-ion battery, and $M^{b}_{nom,cell}$ is the cell nominal mass of battery *b*. The equation for specific energy $E^{b,k}$ for each battery *b* is calculated by charge rates (*C*, amp – hour) times voltage (*V*).

Specific power, $p^{b,k}$, is generated by Li-ion battery and it is calculated by multiplying current (I, amps) and voltage (V). The equation for specific power of battery *b*, cell sample *k*, and applied C-rate *x* is computed as follows:

$$p^{b,k}(x, T_{amb}) = \frac{p^{b,k}(x, T_{amb})}{M^{b}_{nom,cell}}$$

$$p^{b,k} = \frac{I * V}{cell\ mass}$$

4. Analysis

In the research and paper, temperature refers to thermal temperature generated by the battery cell. Also, test time refers to time duration by each individual trial *k* from the experiment. The relationship between the temperature (thermal energy created by the Li-ion battery) and discharge rates is shown by three plots below. The plots are specific energy [*Wh/kg*] – C-rate [*1/h*] for each lithium battery type. Also, enhanced ragone plot (ERp) demonstrates the relationship between specific energy [*Wh/kg*] – specific power [*W/kg*], which demonstrates which battery type shows the best performance and efficiency over test time.

4.1 Specific Energy [Ah] vs C-rate [1/h] for LFP, NCA, NMC

From this graph (figure 1), we can interpret that there is a drastic drop when the charging rate is one-hour. With this increased jump in charging rate, heat will have to be kept in mind as a factor for this drop. The extra heat provided to charge the battery in 20 hours to one hour could alter the difference in performance. The battery gives worse energy as the charging rate increases. However, after charging rate increases past 1C, the energy provided is similar. LFP battery performance in delivering specific energy was the worst at under 0.2 Wh/k*g* in the 5C condition.

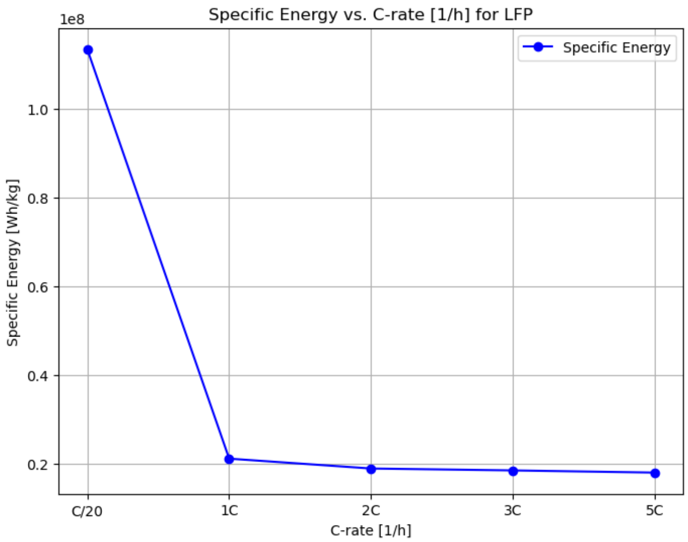

Figure 1. C-rate [1/h] vs Specific Energy [Wh/kg] plot of LFP

The NCA battery (figure 2) follows the same trend as the LFP battery (figure 1). There is a drastic change in specific energy to be delivered between charging rate of C/20 and 1C but that tapers as charging rate increases. Compared to the LFP battery, there is more of a change in energy being delivered as C increases. However, there is still an extreme change between C/20 and 1C. At 5C, both batteries end up being around the same energy being delivered. The highest specific energy given at C/20 at 2.25 Wh/kg compared to the other batteries.

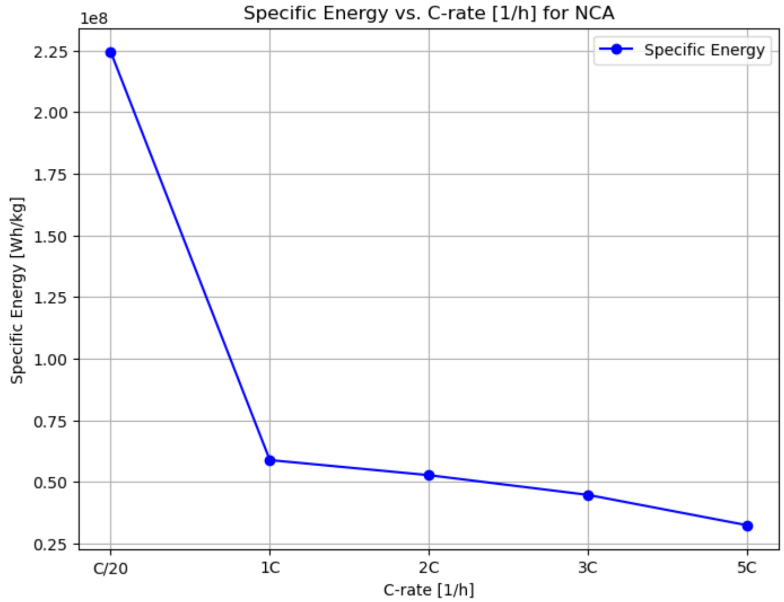

Figure 2. C-rate [1/h] vs Specific Energy [Wh/kg] plot of NCA

For figure 3, NMC follows specific energy decreasing as charging rate increases. NMC follows a similar pattern as NCA and the ending specific energy in relation to 5C. While it is still unclear how temperature is related, it could be a significant factor in decreasing specific energy, or battery capacity, since more heat is generated as charging rate increases.

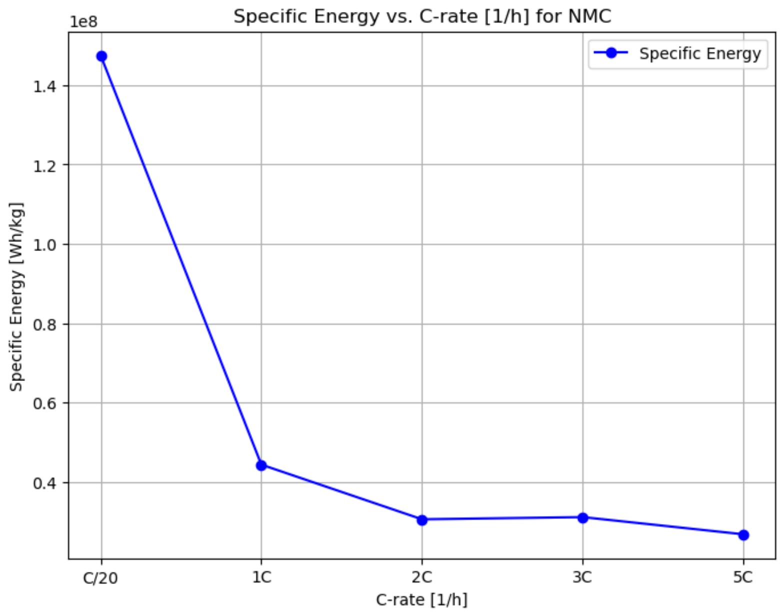

Figure 3. C-rate [1/h] vs Specific Energy [Wh/kg] plot of NMC

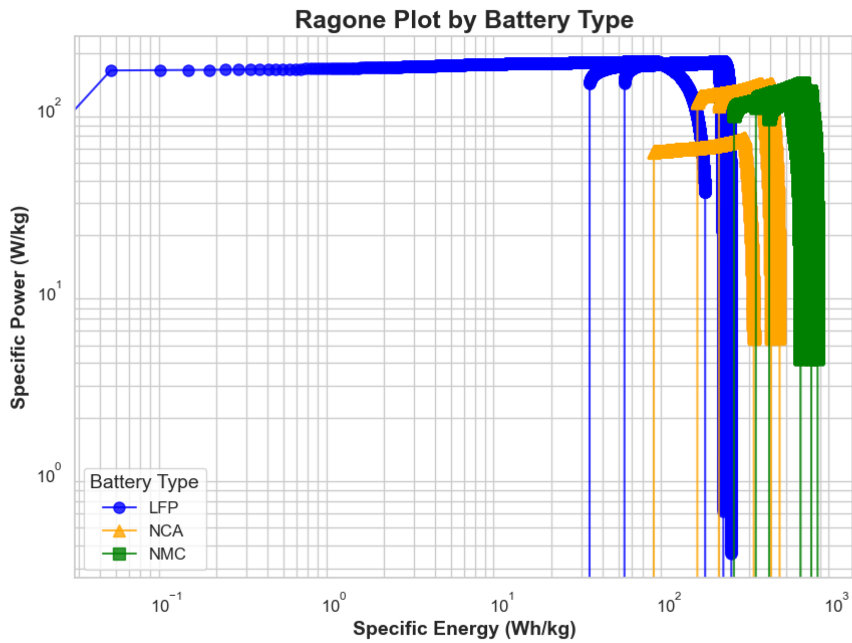

Figure 4. ERp of Specific Power [W/kg] vs Specific Energy [Wh/kg] for LFP, NCA, and NMC at 1C, k1

This Enhanced Ragone plot demonstrates that the NMC battery had the best performance in specific energy (Figure 4). LFP have however, have higher specific power but lower specific energy compared to its counterparts. This means that while LFP's maintain power, they have less capacity for energy storage. There is a wide range of specific energy for LFPs, meaning there is a lot of variability. NCA and NMC are much more stable in terms of specific energy. It offers more specific energy than LFP and less specific power. They are in the middle between LFP and NMC. NMC has the highest specific energy as it is the furthest right. Its specific power is less than LFP though.

4.2 Linear and Polynomial Regression

Our group conducted the regression analysis in two types: linear regression and polynomial regression. In both regressions, we will consider the Surface Temperature (℃) as the response variable, and Current (*I*), Voltage (*V*) and C-rate (charge rate, *1/h*) predictors. According to the correlation heatmap (presented in the Jupyter notebook), the feature which has the strongest correlational relationship with the response variable is current, with a positive correlation of 0.54, while the other independent variables has much less correlational relationship, at –0.16 and 0.20 respectively. The correlation between independent variables and response variable will later be justified further in the linear and polynomial regression.

The linear regression is conducted with the Surface Temperature (℃) be the response variable, whereas Current (*I*), Voltage (*V*) and C-rate (charge rate, *1/h*) be the independent variable. The intercept coefficient in this linear regression is 26.4647, which indicates that if all independent variables equal zero, the surface temperature will be equal to 26.4647 Celsius degree. In this case, and in most cases of linear regression, the intercept coefficient acts as a bias term, and regarding the real-world feasibility, the circumstances of all predictors equal zero does not exist. The slope coefficients of Current (*I*), Voltage (*V*) and C-rate (*1/h*) are -0.8279, -0.4098, and 0.3689 respectively. It indicates that as the predictors increase by 1 unit, the response variable will change within the values of the slope coefficients respectively, holding other variables constant. The R-squared is about 0.35, which is relatively low, indicating that only 35% of the variation of surface temperature is explained by the variation of current, voltage, and C-rate. Additionally, the Mean Square Error is equal to 4.22, which is relatively small considering the range of surface temperature from 5 to 75 Celsius Degree.

On the other hand, the polynomial regression is conducted with the same response variable and independent variables. The intercept coefficient in the polynomial regression is 23.039, which is not a significant change compared to its linear counterpart. The R-squared increased to 0.61, while the Mean Square Error reduced to 2.5, which indicates that the polynomial regression can explain the relationship between features better than the linear regression. To improve the models, we attempt on three different models: Lasso, Ridge and RFE (Recursive Feature Elimination). The results show that Ridge has the same R^2 and MSE with the original polynomial, and both are the best explanatory model, compared to Lasso and RFE. Thus, we decided to keep the original polynomial model.

The polynomial regression analysis is also applied to each of the battery types (LFP, NMC and NCA). The impact of current (*I*), voltage (*V*), and C-rate (*1/h*) varies significantly across battery types (LFP, NCA, and NMC). Current has the strongest negative impact on NCA, followed by NMC, with LFP being the least affected; NCA also exhibits the most severe non-linear effects and the strongest positive interaction with voltage. Voltage decreases output for LFP but significantly boosts output for NCA and NMC, with NCA showing the highest sensitivity, though it suffers from diminishing returns at high levels. LFP experiences positive quadratic returns from voltage, while NMC exhibits moderate diminishing returns. C-rate positively influences all battery types, with NCA being the most sensitive, followed by LFP and NMC. However, NMC suffers the largest diminishing returns, especially at higher C-rate values, while LFP shows the smallest interaction effects overall. In summary, NCA is the most responsive to these features and their interactions, making it highly efficient but more prone to performance declines under extreme conditions, while LFP is the least sensitive, offering greater stability.

By the final polynomial results, which R^2 is equal to 0.61, suggesting that 61% of variations of surface temperature can be explained by the variations of predictor variables, we have the following equation of surface temperature, T:

T = 12.7963 - 6.1898*I + 6.5080*V+ 1.4755*C + 8.553*10^{-5} * (test time) + 0.0219 x I^2 + 1.6952*(I *V) + 0.2643 *(I *C) + -3.983$*$ 10^{-5}*(I *test time) - 0.8460 * V^2 - 0.3802 *(V * C) - 2.852$*$ 10^{-5} * (V * test time) - 0.0841 * C^2 + 3.762*10^{-5}*(C *test time) -7.176$*$ 10^{-11}*($test\ time^2$).

(V x C) is Energy. We shift Surface Temperature to the right-hand side of the equation and Voltage x C-rate to the left-hand side. Thus, we have:

0.3802 * E = 12.7963 - 6.1898*I+ 6.5080*V+ 1.4755*C + 8.553*10^{-5}*(test time) + 0.0219* I^2 + 1.6952 *I*V + 0.2643*I *C + -3.983*10^{-5}*I*(test time) - 0.8460 x V^2 - 2.85210^{-5}*V* (test time) - 0.0841 x C^2 + 3.762*10^{-5}*C *(test time) -7.176$*$ 10^{-11} x $test\ time^2$ + T

Multiplying both sides with 2.6301946344, we have:

$E = 2.6302\ \{12.796 - 6.1898 * I + 6.5080 * V + 1.4755 * C + 8.55 * 10^{-5} * (test\ time) + 0.0219 * I^2 + 1.6952 * I * V + 0.2643 * (I * C) - 3.983 * 10^{-5} * (I * test\ time) - 0.8460 * V^2 - 2.862 * 10^{-5} * (V * test\ time) - 0.0841 * C^2 + 3.762 * 10^{-5} * (C * test\ time) - 7.176 * 10^{-11} * (test\ time)^2 + T\}$.

This is the final equation for Energy, including four different predictors, such as Current, Voltage, C-rate, Test Time, or Energy = f(Current, Voltage, C-rate, Temperature, Test time).

5. Conclusion

This paper investigates the batteries NCA, NMC, and LFP under various temperatures in specific energy, specific powers, surface temperature, current, and voltage. After evaluating the data with linear regression, the formula turns out to be:

$E = 2.6302\ \{12.796 - 6.1898 * I + 6.5080 * V + 1.4755 * C + 8.55 * 10^{-5} * (test\ time) + 0.0219 * I^2 + 1.6952 * I * V + 0.2643 * I * C - 3.983 * 10^{-5} * I * (test\ time) - 0.8460 * V^2 - 2.862 * 10^{-5} * V * (test\ time) - 0.0841 * C^2 + 3.762 * 10^{-5} * C * (test\ time) - 7.176 * 10^{-11} * (test\ time)^2 + T\}$.

This polynomial regression shows surface temperature of the battery is directly contributes to E, with a linear term, +T. This indicates that an increase in surface temperature will increase the energy, all else being constant. However, the impact of surface temperature is additive and does not interact with other variables. Thus, we observe that surface temperature is directly proportional to the energy, while surface temperature maintains independent to other factors including current, charge rate, voltage, and test time.

In the enhanced ragone plot, the NMC battery has the delivers the highest specific energy which means that it can last longer than its counterparts. However, the LFP battery had high specific power but variations in between how much specific energy could be stored. The NCA battery is a balance between the NMC and LFP batteries, delivering moderate specific energy and moderate specific power. This is further backed up by the specific energy graphs in relation to each battery type.

Some difficulties that we encountered were mistakenly calculating our target variable with our features. This means that our polynomial regression and linear regressions had 1.0 R^2 and very low mean squared errors. We also had difficulty using the variables and which equations to use to find the relationship between temperature and energy. By choosing Temperature as the response variable, we can observe the linear relationship and polynomial relationship between Temperature and other predictor variables, and thus, figure out the equation for Energy, or (*V* x *C*). We also attempt to improve the model performance by using Lasso, Ridge, and REF, however, the original polynomial regression model yields the best model performance with R^2 = 0.61.

6. Acknowledgements

We would like to thank Prof. Wang Lulu for the support throughout the semester for DATA300 and final project.

7. Contribution

HyunYoung is a project leader, contributed to data processing (classifiers and training scripts), and data analysis (linear/polynomial regression). Olivia contributed to data collection and algorithm research (data visualization, analysis of data visualization). Minh contributed to data analysis (L2 regularization and linear/polynomial regression).

8. Code

The repository for this project is available at https://github.com/oliviapetronio/Li-ion-batteries-and-Temperatures/tree/main

Resources/References

[1] University of Washington - Clean Energy Institute. https://www.cei.washington.edu/research/energy-storage/lithium-ion-battery/. (accessed Dec. 14, 2024), Lithium-ion battery.

[2] Catenaro, E., Rizzo, M.D., & Onori, S. (2021). "Experimental analysis and analytical modeling of Enhanced-Ragone plot". *Applied Energy*. 291. https://doi.org/10.1016/j.apenergy.2021.116473

[3] *Ltd, S. I. P. (2023, October 21).* "Understanding lithium-ion battery charging: Voltage and current dynamics". *LinkedIn. https://www.linkedin.com/pulse/understanding-lithium-ion-battery-charging-voltage-current-vwhmf/*

[4] *Metrohm Autolab.* "Galvanostatic charge-discharge of a Li-ion battery with Autolab". *https://www.metrohm.com/en_us/applications/application-notes/autolab-applikationen-anautolab/an-bat-002.html*

[5] *Licht, F., Davis, M. A., & Andreas, H. A. (2020).* "Charge redistribution and electrode history impact galvanostatic charging/discharging and associated figures of merit". *Journal of Power Sources, 446, 227354. https://doi.org/10.1016/j.jpowsour.2019.227354*

[6] Catenaro, E., & Onori, S. (2021). Experimental data of lithium-ion batteries under galvanostatic discharge tests at different rates and temperatures of operation. Data in Brief, 35, 106894. https://doi.org/10.1016/j.dib.2021.106894