Capstone Project Vikram Haravu

Udacity Machine Learning Engineer Nanodegree Feb 2018

Contents

[I. Definition 2](#_Toc508121029)

[Project Overview 2](#_Toc508121030)

[Problem Statement 2](#_Toc508121031)

[Metrics 2](#_Toc508121032)

[II. Analysis 4](#_Toc508121033)

[Data Exploration 4](#_Toc508121034)

[Exploratory Visualization 5](#_Toc508121035)

[Algorithms and Techniques 6](#_Toc508121036)

[Benchmark 7](#_Toc508121037)

[III. Methodology 8](#_Toc508121038)

[Data Preprocessing 8](#_Toc508121039)

[Implementation 10](#_Toc508121040)

[Refinement 11](#_Toc508121041)

[IV. Results 12](#_Toc508121042)

[Model Evaluation and Validation 12](#_Toc508121043)

[Justification 12](#_Toc508121044)

[V. Conclusion 14](#_Toc508121045)

[Free-Form Visualization 14](#_Toc508121046)

[Reflection 14](#_Toc508121047)

[Improvement 15](#_Toc508121048)

# Definition

## Project Overview

Rechargeable batteries are a very important component of several daily-use gadgets and EV automobiles too. Demand for Lithium batteries has grown tremendously in the last decade and is expected to continue growing in the foreseeable future. As a result, electronics suppliers and OEMs invest hugely on Battery Management Systems (BMS). Central to the BMS is the function of estimating the State of Charge (SOC) of a given battery. This battery parameter called SoC is directly related to the energy stored inside the battery at any given point of time.

In this project, I developed a Machine Learning model to predict the SoC of a rechargeable battery. It is a regression model since SoC lies between 0.0 (empty battery) and 1.0 (fully charged).

Dedicated and expensive hardware is designed into mobile device ICs to perform the SoC estimation function. I explore an ML method to solve this problem. This project was inspired by this IEEE publication (add link).

## Problem Statement

The following strategy is followed:

1. Generate data using hardware (RTL) simulations. This is a large task by itself. Briefly, a SystemVerilog testbench is built around SoC estimation hardware. The testbench provides stimulus and observes the functioning of the hardware and writes out data files in CSV format.
2. Normalize and Scale the data – since the features/target are different scales of magnitude and units, it is good practice to normalize and standardize them for optimal algorithm performance.
3. Develop a simple benchmark model, to further assess the problem.
4. Develop a model that will better model non-linearities, as measured by metric scores.

For the project, I will address one phase of the BMS/SoC problem – namely, making accurate prediction of the battery’s “*first SoC*” – when a device is powered-on via a battery for the first time. The continuous estimation of SoC during normal operation of a device is critically dependent on accurate estimation of the first SoC. That is because, in the event the first SoC is inaccurate, continuous SoC estimation cannot converge to the correct SoC.

## Metrics

Two metrics will be used in this project – R2 Coefficient of Determination, and MSE Mean Squared Error.

R2 is the square of the correlation (r) between predicted scores and actual scores. It ranges from 0.0 to 1.0. A model with an R2=0.0 (worst score) is a model that always predicts the mean of the target variable. On the other hand, a model with an R2=1.0 (best score) is one that perfectly predicts the target, in other words y\_pred = y\_test.

MSE gives an indication of how close the predicted values are to the real values. MSE is somewhat subjective (data range dependent) in the sense that it doesn’t lie in a fixed range like R2; a smaller MSE is obviously better. Given the mean of the target will be close to 0.0, and the range is approximately -1.0 to +1.0, an MSE of close to 0.0 (i.e., farthest away from 1.0) is a good value to achieve.

Keras models and optimizers will fundamentally attempt to minimize a loss function. MSE is chosen as that loss function in this project.

# Analysis

## Data Exploration

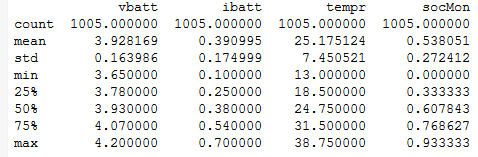
There are only a few proxies/features that are directly measurable and can be thought of as good indicators of the energy stored inside a rechargeable battery. These features are the voltage measured at the battery terminals (vbatt, volts), the current drawn from (or supplied to) a battery (ibatt, amps) and the operating temperature (tempr, Celsius).

**Fig. 1:** These three features and the target variable (socMon) are generated using hardware simulations.

|  |  |
| --- | --- |
|  |  |

The data collected from simulations is shown above. The raw data (CSV file format) is on the left. The raw data needs some minor post processing (such as hex to decimal conversion, divide by 255, etc) to result in actual v/i/tempr/SoC values shown on the right. I chose to not do this raw to actual conversions in the SystemVerilog simulation domain, because it is much simpler to do so with Python.

**Fig. 2:** Here are statistical characteristics of the features and target:



A note on outliers: The simulations to collect data were run under a controlled set of input stimulus. For instance, the hardware module under simulation is spec’d to operate to a max temperature of 39C. The specific battery being acted upon (for SoC est) is rated between 3.6V to 4.2V. These specifications were adhered to when developing simulation stimulus. So, I chose not to remove outliers.

## Exploratory Visualization

**Fig. 3:** In the figure below, I have plotted each feature on individual x-axes with the target socMon on y-axis. One can see the strong or subtle influence of each feature on the target. It also shows the direct or inverse relation between each feature and target.

|  |
| --- |
|  |

**Fig. 4:** In the plots below, a subset of the data is used. All data-points which satisfy the following condition are removed: (ibatt < 0.3 & ibatt > 0.5). It is seen that while the relation with vbatt and tempr are similar to the full-data plots, the relation with ibatt has become even more subtle. This gives an indication of the nature of non-linearity inherent to the SoC estimation problem.

|  |
| --- |
|  |

**Note:** the subset used for this plot is only for initial exploratory purposes and not for model development.

## Algorithms and Techniques

A Deep Learning model is proposed for this problem. It will be a Fully Connected MLP type. Given the size of the dataset and that MLPs are known to work well for regression problems, it is decided to go with this type.

The following hyperparameters can be tuned during the search for the best model:

* Epochs
* Batch size – the number of data-points to use for one feedforward-backprop step in each epoch. Example: Say X\_train=1000, and batch\_size=10, there will be 1000/10=100 ff-bp iterations (of 10 samples) per epoch to arrive at the weights/bias set. For a given problem, it is an experimental (hyperparam search) process to arrive at an optimal batch\_size.
* Optimizer – Stochastic Gradient Descent (sgd)
* Learning rate and Decay
* Momentum – to avoid local minima

Tuning of the Fully Connected network architecture:

* A wide vs. deep network
* Dropout – to avoid overfitting
* Weights initialization
* Activation: relu will be used because it is a continuous function; and given we are solving a regression problem, a continuous activation function is appropriate.

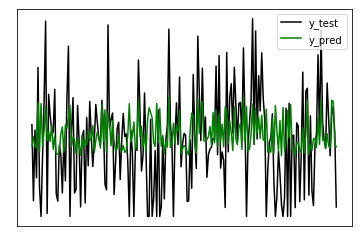
## Benchmark

A linear model that uses the most correlated feature from Fig. 3 and Fig. 4 will be developed to model the target. The voltage at the terminals of a battery is the best indicator of the energy stored inside the battery. Scikit’s linear\_model.LinearRegression() was used for the benchmark.

This model resulted in the following metrics-

* + MSE = 0.961 and
  + R2 = 0.066

**Fig. 5** Benchmark model performance visualization



**Note 1:** Pre-processed data was chosen to develop the benchmark. Details about pre-processing are discussed next.

**Note 2:** Comparison of the benchmark and best model is discussed in the Results section.

# Methodology

## Data Preprocessing

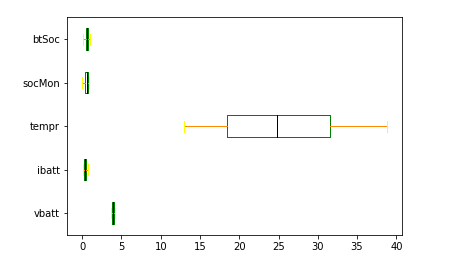
The distribution characteristics of the data is discussed below-

Most of the data in our dataset is normally distributed – the median and mean are close to each other. This can typically be observed by plotting boxplots of the data.

**Fig. 6**: Data distribution

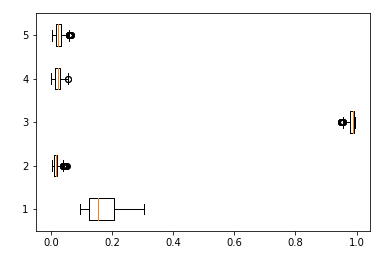
|  |  |
| --- | --- |
|  |  |
|  |  |

**Fig. 7**: Boxplots of all features and target (in a single view) – it shows varying orders of magnitude.



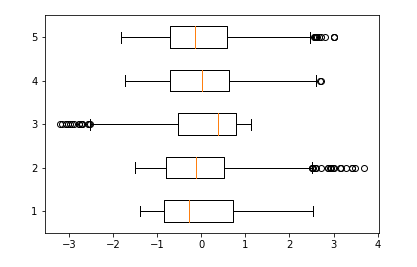
It is appropriate to normalize numerical data that lie in different magnitude ranges. ML algorithms perform better on data thus pre-processed.

**Fig. 8**: Boxplots of all features and target after normalization – all lie between 0.0 and 1.0



**Scaling**: this is also an important pre-processing step. This step will achieve the data to have a mean of approximately zero. An optimizer like SGD can converge to the minimum, more effectively on norm’d and scaled data.

**Fig. 8**: Boxplots of all features and target after normalization and scaling:



The normalized and scaled data is then split into training and testing set (20%). The resulting shapes as shown below:

X\_train (804, 3)

y\_train (804,)

X\_test (201, 3)

y\_test (201,)

## Implementation

A Keras sequential() model with three hidden layers in addition to the input layer and output layer is built. The rectified linear unit “relu” activation is used for all layers, except the final layer. That is because, given the regression problem, the predicted output should be a numerical value without any transform. Dropout is added to every layer to prevent over-fitting. With dropout, a percentage of nodes will be inactive during feedforward-backprop steps of training – thus avoiding the possibility of a few nodes “controlling” the network.

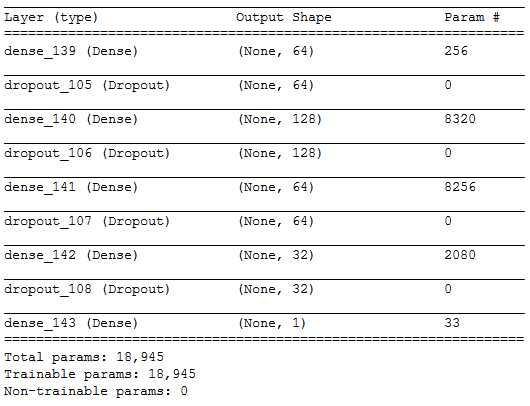
It is valuable to initialize the weights and biases prior to training. The benefit can be understood by acknowledging the fact that if the weights are initialized to a very small value, then the signal gets far too attenuated as it passes through each layer, eventually becoming negligible. On the other hand, too large weights will cause the signal to blow up in magnitude. I have chosen weights initialization RandomNormal() with a mean=0.0.

The loss to be optimized is MSE and the optimizer is SGD.

Training is checkpointed – to save only the best model weights as training proceeds. The weights are stored in the saved\_models/ directory. These weights then are loaded after training is completed, to evaluate the model performance.

The training data is further split to result in validation data of size 10%. This validation data slice is unseen data during each training epoch. At the end of an epoch, performance (val\_loss) of the model is evaluated on the validation data to conclude if an improved model has resulted in this epoch.

**Fig. 9:** Model summary



**Note on the # parameters**: as an example, the param# for the input layer is determined as follows:

* + ([wt=64] x 3) + [bias=64] = 256. (64 is the number of nodes, 3 is the feature shape)

## Refinement

Inferences of the model from the previous section resulted in a marked improvement in metrics from that achieved by the benchmark.

|  |  |  |
| --- | --- | --- |
|  | **Benchmark model** | **Initial MLP** |
| **MSE** | 0.961 | 0.340 |
| **R2** | 0.066 | 0.669 |

For refining the model, the following best practices (industry standard) are applied:

**Learning rate:** “lr” is a small value, default = 0.01 for Keras’s optimizers. This is used as a multiplier to figure out by how much to correct weights and biases during training epochs. The related decay parameter defaults to 0.0; which means that the learning rate is constant for all epochs. By allowing the learning rate to decay with time (epoch #), smaller changes to the parameters are made during mature training epochs, thus avoiding overcorrecting.

The following is used: lr=0.01, decay=1e-6

**Batch\_size:** This is changed from the default of None to 8, which resulted in better performance than an increase to 64.

**Num epochs**: Increased to 500. Val\_loss improvement was observed up until epoch #440

# Results

## Model Evaluation and Validation

The refinement discussed above resulted in substantial performance improvement from that of the initial model:

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Benchmark model** | **Initial MLP** | **Refined model** |
| **MSE** | 0.961 | 0.340 | 0.054 |
| **R2** | 0.066 | 0.669 | 0.948 |

A wide MLP network with just one 256-wide hidden layer was developed. The results of the wide vs. deep model were comparable to each other.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Benchmark model** | **Initial MLP** | **Refined model** | **Wide model** |
| **MSE** | 0.961 | 0.139 | 0.054 | 0.061 |
| **R2** | 0.066 | 0.865 | 0.948 | 0.941 |

## Justification

To have an appreciation for the robustness of the final solution, the following two points are to be noted:

1. By going with two metrics for this project, namely, MSE and R2, a high standard was set. MSE is a metric that is data dependent, and the R2 score gives an objective score to the model. The final model high scores on both metrics.
2. The visualization below suggests that non-linearities between the 3 features have been modeled well.

**Fig. 10:** y\_test vs. y\_pred for the various models

|  |
| --- |
| Benchmark model |
|  |

|  |
| --- |
| Initial DL model |
|  |

|  |
| --- |
| Refined final model |
|  |

# Conclusion

## Free-Form Visualization

In the plots below, I compare the training history of the initial model with the refined model.

**Fig. 11**: Training history comparison

|  |
| --- |
| Initial model |
|  |

|  |
| --- |
| Tuned model |
|  |

The val\_loss of the tuned model begins to converge towards training loss at an early stage. This indicates a better use of epochs. The reduced batch size for the refined model is the enabler for this improvement.

With the initial model, during later epochs, there seems to be overfitting happening frequently. This can be seen by the large val\_losses around the 275+ epochs. On the other hand, in the tuned model logs, this issue isn’t so pronounced. It is likely that the small (dampened) learning rate applied during the time of the late epochs is instrumental in controlling overfitting.

## Reflection

## Improvement