RLU estimation

Thai Nguyen

1 Introduction

1.1 Problem formulation

1.2 C-MAPSS dataset

This dataset was generated with the C-MAPSS simulator. C-MAPSS stands for "Commercial Modular Aero-Propulsion System Simulation" and it is a tool for the simulation of realistic large commercial turbofan engine data. The C-MAPSS dataset contains 4 sub-datasets generated under different operating and fault conditions which are further divided into training and test sets. A summarization of this dataset is given as belows:

Dataset	FD001	FD002	FD003	FD004
Number of training trajectories	100	260	100	249
Number of test trajectories	100	259	100	248
Number of operating conditions				
Number of fault modes				

Table 1: C-MAPSS dataset

The data are provided as a zip-compressed text file with 26 columns of numbers, separated by spaces. Each row in the data is a snapshot of data taken during a single operating time cycle, which includes 26 columns:

- The first column represents the engine ID.
- The second column represents the current operational cycle number.
- Columns from 3 to 5 are the three operational settings that have substantial effects on engine performance.
- Columns from 6 to 26 represent the 21 sensor values.

1.3 Objective

The goal is to predict the number of remaining operational cycles before failure in the test set, i.e., the number of operational cycles after the last cycle that the engine will continue to operate.

2 Data processing

2.1 Piece-wise linear RUL target function

2.2 Data normalization

The goal of normalization is to transform features to be on a similar scale which can improves the performance and training stability of the model. It is often used when data futures are on drastically difference scales.

2.2.1 Max-min normalization

2.2.2 Clipping

2.2.3 Log scaling

2.2.4 Z-score