



Presentation Team Imputation

BY: RAMON, MICHAEL, JESÚS, ALBERT, ADRIEN & JULIËN



Table of contents

- Improvements to pipeline
- Imputation by Regression
- Imputation by Hot Deck
- Neural networks
- Research paper

Improvements to pipeline

- Improved data loading:
 - Multiple features
 - Multiple targets

Load clean data

The upload form only supports files up to 10Mo. For larger files, please upload them directly to JupyterHub and provide a relative link to them in the text input herebelow.

EDIT: customize the dataset configuration.

In [20]: `%run helpers/load_data.ipynb`

Config: factory zero

Explorer

Upload

File path

Uploaded file selector:

054.xlsx

smartMeter

flowHeatSpaceHeatr

Add sheet

Columns type:

☒ features

☐ targets

Sheet:

smartMeter

Columns:

energy_in_low

energy_in_norm

energy_out_low

energy_out_norm

power

Discard sheet

Imputation by regression

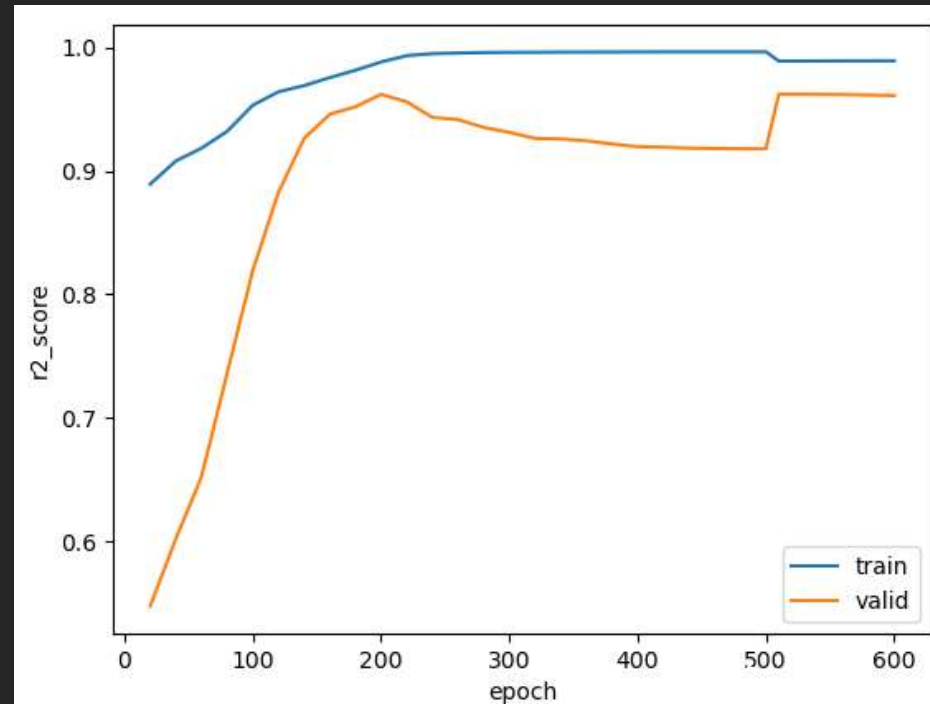
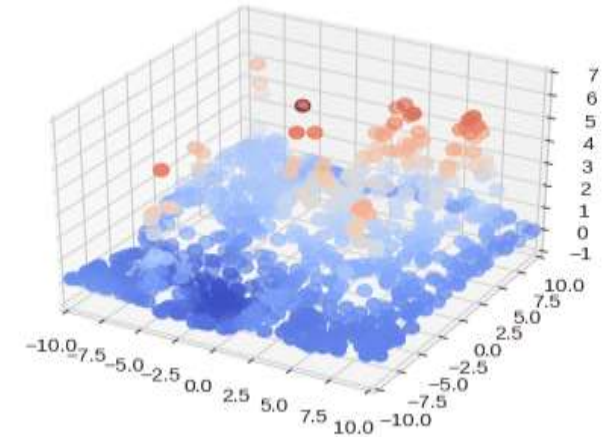
- Deterministic Regression
- Stochastic Regression
- On pipeline no satisfactory results
- Difference
 - Deterministic: fully determined by parameters
 - Stochastic: adds randomness by probability distribution

Imputation by Hot Deck

- What is hot deck?
- Vectorization
- Precision

Neural Networks

- Recurrent Neural Networks practice with lectures data
- Optimizing hyperparameters using genetic algorithm
- 0.85 \rightarrow 0.963 r2_score (40 generations 21,500 genomes)



Research (Paper)

- Introduction
- Methodology
- More in depth research for:
 - Hot Deck Imputation
 - Recurrent Neural Networks
 - Other methods optimized for time series data

What we will be working on?

- Improving Hot Deck
 - Using RNN on houses data
 - Other "easier" models
 - More focus on paper
-
- From data collection to data evaluation
 - No new big methods

Questions & Feedback

- Questions:
 - Are there any other imputation methods for time series data we should try to implement?