INM427 Neural Computing

Supanut Sookkho (MSc Data Science / 230024841)

Yumi Heo (Msc Data Science / 230003122)

A Comparison Between Python and MATLAB for Binary Classification in Predicting Bank Churn

Background of the data set, its format, basic statistics, nn methods  
some background reading about the data set, its format, size and basic statistics,

* We use bank churn data set from Kaggle[1] to compare the process of binary classification.
* 165034 rows × 14 columns

Libraries and Frameworks:

Python offers a rich ecosystem of libraries and frameworks for machine learning and data analysis, such as scikit-learn, TensorFlow, and PyTorch. These libraries provide various algorithms and tools for building and evaluating classification models.

MATLAB provides its own set of tools and libraries for machine learning, including the Statistics and Machine Learning Toolbox and the Neural Network Toolbox. While these tools are powerful, the range of available algorithms may be more limited compared to Python.

What methods we use for our models

the NECO methods that might have been already applied to it

Consider if there is another method that can be suitable

which other methods might be suitably applicable to your data

Explain modeling

Compare the process and results from python and matlab

“and the NECO methods that might have been already applied to it.

consider which other methods might be suitably applicable to your data. This will require you to read a few papers that have used the data set and the general overview descriptions of the NECO methods studied in INM427.

an initial analysis of the data set including basic statistics and a comparison of the results, including two graphs showing Matlab’s and Python’s training speeds and test set accuracies, and references to the papers you’ve read.”

# Matlab can also select multiple activation functions

# parameter tuning is limited compared to Python. (need to discover it more)

# MATLAB has standardization by default.

# Compare normalization and standardization (Python, Matlab)

# We can also compare model performance with the confusion metric. (F1, precision, recall)

# Speed (For example, with the same set of training data why is MatLab faster than PyTorch)

# Learning rate (Which platform converges faster)

References