

Experiments protocol

Léon Delachaux

1. Overview

- The experiments aim to analyze data from the CSCR_data dataset using various machine learning techniques.
- The experiments are implemented in Python scripts and a Jupyter notebook.

2. Directory Structure

- Root Directory:

T:\Studies\CSCR\code\cscr_stats_ml

- Gitlab:

[papers / CSCR_stats_ML · GitLab \(sib.swiss\)](#)

- Contains:
 - Data file: CSCR_data.json
 - Code files:
 - correlation.py: Computes Pearson correlation between pairs of biomarkers.
 - stat_test.py: Implements t-test and Kolmogorov-Smirnov (K-S) test.
 - stat_model.py: Implements logistic regression statistical model.
 - utils.py: These are utilities functions used inside the three other py files
 - Notebook: experiment_pipeline.ipynb
 - Environment file: ml_paper_env.yml

3. Environment Setup

- Create the environment using `conda env create -f ml_paper_env.yml`.
- Modify the prefix in the YAML file with the desired environment name.
- After setup, activate the environment using `conda activate ml_paper`.

⚠ I had a problem with the seaborn library to display the correlation matrix. If it happens, try to reinstall it with pip (`conda uninstall seaborn, then pip install seaborn`)

4. Running Experiments

- Open and run the notebook `experiment_pipeline.ipynb`.
- Execute all cells in the notebook to obtain results.

5. Results

- Upon running the notebook, the following results will be generated:
 - Correlation matrix
 - T-test results (for single visit data)
 - K-S test results (for longitudinal data)
 - Predictive modeling experiment

6. Figures

- Figures are generated within the notebook, allowing for easy reformatting for publication purposes.