
Human Cognitive Performance:

Regression problem

Context

In this practical work, you will work with Human Cognitive Performance dataset. The goal of your algorithm is to accurately predict three values : 1. Reaction Time ; 2. Memory Test Score ; 3. Cognitive Score (note that the AI_Predicted_Score column is to be discarded). You can find the dataset on Kaggle¹, along a description of the dataset. For this work, you will work with another student.

You can also find some implementation on Kaggle for this dataset, feel free to explore them after the practical work, but refrain from using them during your try: the main point of that practical work is to struggle a bit, as if you are encountering a completely new problem in your company and learn from it.

Part 1: Evaluation Methods Setting

For this part, you can use Scikit-Learn, both students should answer it together.

1. Explore and understand the dataset with your teammate.
2. Determine and implement the metrics you will use to evaluate your models' performances. Besides performances metrics for the prediction, also retrieve the execution time. Define your priorities and explain why.
3. Implement Dummies methods (at least 2) and evaluate their performances.
4. Implement other regression models / baselines (at least 2) and evaluate their performances.
5. What can you observe from the performances results?

¹ <https://www.kaggle.com/datasets/samxsam/human-cognitive-performance-analysis/data>

Part 2: MLP Regression

On this part, one student will use TensorFlow and the other one Pytorch to implement the same answers.

The maximum number of neurons allowed is 100 (you can use less), they can be split in up to 3 hidden layers.

6. Implement and evaluate an MLP Regressor with default values using the metrics defined in part 1. What can be said about the model performance and what can be inferred from it about the problem to solve?
7. With the help of the course, identify several hyper-parameters (excluding regularization techniques) to fine-tune to improve your model. Note the configuration and performances on the Table shared with the class (url on moodle). Report for at least 5 configurations (on top of the default) not previously in the table.
8. For each configuration, experiment with different regularization techniques (dropout, pruning, L1/L2...), note your observations.
9. Based on the reported results, what can we observe about:
 - The influence of the **framework** on the performances and learning.
 - The influence of the **hyper-parameters** on the performances and learning.
 - The influence of the **regularization techniques** on the performances and learning.

Write those observations as comments in your code or a separate text file, they will be discussed in class. Discuss with your teammate for a better understanding of the frameworks differences and the interpretation of the performance metrics.