TASK 1

Brain Age Prediction

You are in the group Whatever consisting of fpalmegi (fpalmegi@student.ethz.ch (mailto:///[u'fpalmegi@student.ethz.ch'])), luciensc (luciensc@student.ethz.ch (mailto:///[u'luciensc@student.ethz.ch'])) and sgiles (sgiles@student.ethz.ch (mailto:///[u'sgiles@student.ethz.ch'])).

1. READ THE TASK DESCRIPTION

2. SUBMIT SOLUTIONS

3. HAND IN FINAL SOLUTION

1. TASK DESCRIPTION

TASK 1: PREDICT THE AGE OF A BRAIN FROM MRI FEATURES

This task is primarily concerned with regression. However, we have perturbed the original MRI features in several ways. You will need to perform outlier detection, feature selection, and other preprocessing to achieve the best result.

DATA DESCRIPTION

Download Data (/static/task1.zip)

The data for this task contains the following files:

- **X_train.csv** - the training features
- **y_train.csv** - the training targets
- **X_test.csv** - the test features (you need to make predictions for these samples)
- **sample.csv** - a sample submission file in the correct format

Each row in X_train.csv is one sample indexed by an id, so the first column contains the id. In addition to the id column, each sample has 833 features:

```
id,x0,x1,...,x833
0,7077.5,926681.1,...,65052.6
...
```

The test set file (X_test.csv) has the same structure:

```
id,x0,x1,...,x833
0,6516.1,979705.6,...,65052.6
...
```

The training targets (age in years) are contained in y_train.csv:

```
id,y
0,75
1,76
2,74
...
```

For your convenience, we further provide a sample submission file (sample.csv):

```
id,y
0,69.8
1,69.8
2,69.8
...
```

Note that, for each prediction you need to include the id of the sample as in X_test.csv.

SUBMISSION FORMAT

For every data instance in the test set, submission files should contain two columns: *id* and *y* where *y* should be a double with your prediction.

The file should contain a header and have the following format:

```
id,y
0,77
1,56
2,89
...
```

Please keep in mind that, as a group, you have a limited number of submissions as stated on the submissions page.

EVALUATION

The evaluation metric for this task is the **Coefficient of Determination (R^2) Score** which ranges from minus infinity to 1.
Given true values y_i, and predicted values $\hat{y}_i$, the formula reads:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}$$

with $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$.

How to compute it in Python:

```
from sklearn.metrics import r2_score
score = r2_score(y, y_pred)
```

⚠ Make sure that you properly hand in the task, otherwise you may obtain zero points for this task.

FREQUENTLY ASKED QUESTIONS

WHICH PROGRAMMING LANGUAGE AM I SUPPOSED TO USE? WHAT TOOLS AM I ALLOWED TO USE?

You are free to choose any programming language and use any software library.

CAN YOU HELP ME SOLVE THE TASK? CAN YOU GIVE ME A HINT?

As the tasks are a graded part of the class, **we cannot help you solve them**. However, we will try to address general aspects during the project tutorials. Moreover, feel free to ask general questions about the course material during or after the exercise sessions.

CAN YOU GIVE ME A DEADLINE EXTENSION?

> ⚠ We can not grant deadline extensions, except in extraordinary cases (e.g. military service). However, we will require official confirmation of your problem (e.g. certificate of illness).

CAN I POST ON PIAZZA AS SOON AS I HAVE A QUESTION?

This is highly discouraged. Instead,

- Read the details of the task thoroughly.
- Review the frequently asked questions.
- If there is another team that solved the task, try again.
- Discuss with your team mates.

If you still consider that you should contact the TAs, you can post a **private** question on Piazza. Remember that collaboration with other teams beyond (general discussions) is prohibited.

WHEN WILL I RECEIVE THE PRIVATE SCORES? AND THE PROJECT GRADES?

We will publish the private scores before the exam the latest.