Californa Socioeconomic Correlations Marking Scheme

Eirik Albrektsen [1]

[1]Department of Electrical and Electronic Engineering
Imperial College London

November 16, 2018

1 Submission

To get access to the test set contact Eirik Albrektsen on slack to get a test set without the values you are trying to predict by 12.00 AM Sunday

Once you have received this you should output your predicted y-values or any findings into a `.csv` file and upload it, together with all the codes used, onto the designated folder, and then zip it:

In particular upload a zip file such that it has the following structure:

- {Team Number}.zip
  - {Team Number}_{Feature Short Name}.csv
  - Short Name can be found in BG_METADATA_2016 under Short_Name
  - {Team Number}_report.pdf

For example, one could submit the following:

- 2.zip
  - 2_B00001e1.csv
  - 2_report.pdf
  - 2_data_preprocessing.ipynb
  - 2_model.py
  - 2_presentation.pptx

2 Guideline

To get access to the test set contact Eirik Albrektsen on slack by 12.00 AM Sunday. He will give yoy a test set without the values you are trying to predict.

A short report is to be submitted atsubmitted onto the Google Forms by 14:00 18 November 2018. Later submissions will not be accepted.

You may submit your report in the form of PDF, HTML or web link. If you are submitting a website, please include the link in a `.txt` file.

3 Marking scheme

We will judge you by the final quality of your report: rationale of your model(s) and the clarity of your explanation in the final report are important.

The following marking scheme for report is adopted from the Student Handbook in MSc Bioinformatics and Theoretical Systems Biology at Imperial College London

Understanding of socioeconmic concepts (10%):

- Complete understanding of the material to be presented, in particular the socioeconomic concepts relating relevant to this task
- Preparatory work including sufficient background reading in socioeconomics and machine learning

Presentation (15%):

- Clear and logical organisation of the material following good academic practices such as proper use of references and scientific writing style
- Present the material in a concise manner and appropriate use of tables and diagram, adhering to the word limits for the report and abstracts

Potential (15%):

- Demonstrate the impact and implications of your results.
- Present how your results can be used in the real world, either by industry, academics, or public services

Originality (20%):

- Demonstrating creativity in reasoning why one variable is more interresting than another.
- Application of existing methods to new datasets and justification of methods.

Methodology and Results (40%):

- Excellent use of appropriate machine learning algorithms and mathematical models, in particular to the merit and limitations of different algorithms and how these methods are used suitably to address the biological problem presented
- Good use of programming techniques, including but not limited to choice of programming languages and packages, data structures and algorithms, software development practices
- Justification of results and observations from data using socioeconomical and mathematical knowledge and other suitable means