Project 2020

Fundamentals of Data Analysis

Due: last commit on or before January 8th, 2021

These are the instructions for the Project assessment for Fundamentals of Data Analysis in 2020. The assessment is worth 50% of the marks for the module. Please read the *Using git for assessments* [2] document on the Moodle page which applies here. As always, you must also follow the code of student conduct and the policy on plagiarism [1].

Instructions

In this project you must perform and explain simple linear regression using Python on the `powerproduction` dataset available on Moodle. The goal is to accurately predict wind turbine `power` output from wind `speed` values using the data set as a basis. Your submission must be in the form of a git repository containing, at a minimum, the following items:

1. Jupyter notebook that performs simple linear regression on the data set.
2. In that notebook, an explanation of your regression and an analysis of its accuracy.
3. Standard items in a git repository such as a README.

To enhance your submission, you might consider comparing simple linear regression to other types of regression on this data set. Rest assured, all the above concepts will be explored in lecture videos and other materials in the coming semester.

Marking scheme

The following marking scheme will be used to mark your submission out of 100%, which will then be scaled to 50%. The examiner's overall impression of your submission may influence marks in each individual component. It is important that your submission provides direct evidence of each of the items listed in each category. For instance, your commit history should demonstrate and provide evidence that you had a pragmatic attitude to completing the assessment. Likewise, your submission should have references in it to demonstrate that you considered the literature and the work of others.

25%	**Research**	Evidence of research performed on topic; submission based on referenced literature, particularly academic literature; evidence of understanding of the documentation for any software or libraries used.
25%	**Development**	Environment can be set up as described; code works without tweaking and as described; code is efficient, clean, and clear; evidence of consideration of standards and conventions appropriate to code of this kind.
25%	**Consistency**	Evidence of planning and project management; pragmatic attitude to work as evidenced by well-considered commit history; commits are of a reasonable size; consideration of how commit history will be perceived by others.
25%	**Documentation**	Clear documentation of how to create an environment in which any code will run, how to prepare the code for running, how to run the code including setting any options or flags, and what to expect upon running the code. Concise descriptions of code in comments and README.

References

[1] GMIT, "Quality assurance framework," https://www.gmit.ie/general/quality-assurance-framework.

[2] I. McLoughlin, "Using git for assessments," https://github.com/ianmcloughlin/using-git-for-assessments/.