PHYS 243

Mid-term Project

Deadline: May 16, 2021, 11:59 pm

Data sets:

You can choose any of the following data-sets:

1. Climate Model Simulation Crashes Data Set
2. Acute Inflammations Data Set
3. Teaching Assistant Evaluation Data Set
4. Wine Quality Data Set

You can find the information on each data set from the link below, which contains the data itself, what each columns represents and additional information on the data sets.

Link to data sets directory

Expectation

You can use any python package that you want, however, you should explain what each function/object is doing and as well as their parameters and why you have chosen those for the problem at hand.

1. Reading the dataset and exporting it into your workflow.
2. Understanding and explaining the data set, what are the different attributes of each data point.
3. Processing, and cleaning up the data as needed.
4. Dividing your data into a training and test set.
5. Choosing the relevant algorithm and why you think it is better fit for this data set.
6. Writing a python code to perform learning. (You can reuse every code from the lectures)
7. Evaluating your model's performance.
8. Making sure your results does not depend on the choice of parameters.

You should explain every step clearly