MASSACHUSETTS INSTITUTE OF TECHNOLOGY

DEPARTMENT OF MECHANICAL ENGINEERING

2.s986 *Applications of Machine Learning in Engineering Science*
Fall Term 2020

PSet #1 | Distributed September 3, 2020 | Due September 10, 2020

This PSet consists of three problems designed to introduce the learner to workflow management. For all problems below, show your work and assumptions for full credit.

Problem 1: For the handwritten digit recognition problem presented in Lecture 1 (available in the footnote[1]):
   a) Write a block diagram to map the workflow (from data curation to machine learning prediction).
   b) Estimate (quantitatively) the time required for each step, and qualitatively describe the complexity of each step.

Problem 2: Repeat the exercise above, but for a workflow in your experimental (or theoretical / simulation) research. If you are not an active researcher, pick a problem similar to research, *e.g.*, heart disease classification. (Hint: You can also pick a dataset from Kaggle[2])

Problem 3:
   a) Back to the handwritten number dataset, select different algorithms (using the code provided[1]) and tune the hyperparameters to improve the prediction accuracy for handwritten digit dataset. Find one combination of algorithm, and hyperparameters, that offers an improved predictive accuracy over the baseline case presented in the code (just above 91%). Provide the algorithm name and hyperparameters in your solution.
   b) What you have just done, is tuned a "black-box" machine-learning algorithm, perhaps without any underlying knowledge of how it works. Describe what you think might be potential drawbacks of this approach — and when it might be necessary.

Please submit your report and code in a zip file. You are encouraged to set up an anaconda environment and run the code locally. You can refer to footnote[3] for setting up the anaconda environment.

---

[1] https://github.com/PV-Lab/2s986_class/blob/master/Week1/Classification_demo.ipynb
[2] https://www.kaggle.com/tags/classification
[3] https://github.com/PV-Lab/2s986_class/