

Assignment 2: Machine Learning
Prepared by: Miguel V. Martin, PhD, PEng
Due date: October 30 2022, by 11:59 PM

Instructions: This is an individual assignment; collaboration is not allowed. Submit the solutions using Canvas' submission system. Please follow the submission instructions specific for each question.

Notes on the **Academic Integrity Policy** (<https://usgc.ontariotechu.ca/policy/policy-library/policies/academic/academic-integrity-policy.php>):

- a) If you reuse code written by someone else: (a) Make sure you credit the source/author of the reused code by clearly stating the origin of that code using comments as appropriate. (b) Beware that you can only receive credits for your own code. (c) Failing to credit the source/author of the reused code is considered academic misconduct. Please refer to articles 16 through 19.
- b) It is prohibited to disseminate/post this assignment or parts of it. Please refer to Article 20.

1. **[Marks 60%]** Use the UCI Abalone dataset `abalone.data` available at <https://archive.ics.uci.edu/ml/datasets/Abalone>

- a) [Marks: 0%] Split the data into 50% training and 50% validation sets.
- b) [Marks: 20%] Train a poor deep neural network (DNN) model using `keras.models.Sequential` to classify abalones into “male”, “female”, or “infant”, in such a way that the model intentionally overfits the training data. Use 100 epochs. Generate a graph that depicts `loss`, `accuracy`, `val_loss`, and `val_accuracy`.
- c) [Marks: 20%] Improve the model from Part b) by changing some, or all of the following hyperparameters: kernel initializer, activation function, normalization, regularization, optimizer, and learning rate schedule. Use 100 epochs. Generate a graph that depicts `loss`, `accuracy`, `val_loss`, and `val_accuracy`. You may want to use a utility called *Keras Tuner*.
- d) [Marks: 20%] Your code should display each model from b) and c) using `keras.utils.plot_model`.

Submit:

- [1] your code as a Jupyter notebook named `abalone.ipynb`
- [2] the graphs in separate files called `abalone-c.png`, and `abalone-d.png`, and
- [3] both models from Part d) as files called `abalone-db.png` and `abalone-dc.png`

2. [Marks: 40%] Use the Auto MPG dataset `auto-mpg.data`, available at <https://archive.ics.uci.edu/ml/datasets/Auto+MPG>
- [Marks: 0%] Drop the `name` and `origin` attributes.
 - [Marks: 0%] Attribute `horsepower` has 6 missing values. Fill these up with the average horsepower for the respective number of cylinders.
 - [Marks: 5%] The `year` attribute has two digits. Add 1900 so that the year has four digits. For example, in the dataset “70” really means “1970”, so you need to change the value from 70 to 1970.
 - [Marks: 5%] Split the data into 50% training and 50% validation sets.
 - [Marks: 15%] Train a `keras.models.Sequential` DNN that predicts the `mpg` of vehicles.
 - [Marks: 15%] Use your DNN from Part e) to predict the `mpg` of the following vehicles:

vehicle	cylinders	displacement	horsepower	weight	acceleration	year
1*	6	2170	502	3164	4.2	2022
2*	12	6498	730	3472	3.2	2022
3*	8	3902	986	3020	2.5	2022
4*	8	6162	670	3721	2.6	2022
5	4	122	181	2496	8.3	2022
6	6	3232	155	3232	11.5	1969
7	3	598	89	1550	10.1	2022
8	3	900	50	642	5.8	2022
9	4	1189	60	2355	28.1	1964
10	4	201	40	2265	32	1908

*The data in the first four rows resemble the characteristics of modern high-end sports cars.

Submit:

- [1] your model’s predictions by having your program print them but also including them as comments within your code.
- [2] your code in a Jupyter notebook called `auto.ipynb`
- [3] a short paragraph as a comment at the end of your code, providing a rationale for your model’s predictions.