SEIS 764 Artificial Intelligence
Assignment 2
Due: midnight Friday 10/1/21 on Canvas

Individual effort

Building a NN Regression Model

Dataset: The dataset is about the compressive strength of different samples of concrete based on the volumes of the different ingredients that were used to make them. Ingredients include Cement, Blast Furnace Slag, Fly Ash, Water, Superplasticizer, Coarse Aggregate, Fine Aggregate, and Age. These are the features (predictors). The target column is the Strength. If you look at the first sample, you can see that it has 540 cubic meter of cement, 0 cubic meter of blast furnace slag, 0 cubic meter of fly ash, 162 cubic meter of water, 2.5 cubic meter of superplaticizer, 1040 cubic meter of coarse aggregate, 676 cubic meter of fine aggregate. Such a concrete mix which is 28 days old, has a compressive strength of 79.99 MPa.

Create a notebook called **Assign2.ipynb**. Write code for each of the following questions by having a separate cell for every question. Copy the actual question in a markdown cell and right below that you should have a code cell as shown below.

1. Load the dataset and print how many data instances (samples) we have.

```
In [ ]: ## Your code here
```

1. Load the dataset and print how many data instances (samples) we have.
2. Check if there are any missing values in any of the columns.
3. Split the data into the features (i.e. X) and the target (i.e. y). Print first few rows of X and y.
4. Normalize all columns of X using standardization. That is, subtracting the mean and dividing by the standard deviation.
5. Build a simple NN Regression model that has one Dense unit. Specify **mean_squared_error** for the loss and **adam** for optimizer. Allocate 30% of the data for validation. Run it for 100 epochs. Plot the loss plot for training and validation.
6. Build a simple NN Regression model that has one Dense unit. Specify **mean_squared_error** for the loss and **sgd** for optimizer. Allocate 30% of the data for validation. Run it for 100 epochs. Plot the loss plot for training and validation.
7. Build a simple NN Regression model that has one Dense unit. Specify **mean_squared_error** for the loss and **rmsprop** for optimizer. Allocate 30% of the data for validation. Run it for 100 epochs. Plot the loss plot for training and validation.
8. Which network performed the best? Do you have ideas on how the performance can be improved? Answer this question in the markdown cell in the notebook.

Note: You can read more about the optimizers at https://keras.io/api/#optimizers

Submission:

- Make sure each of the cells have been run with the output shown right below. Now, export the notebook as .html file.
- Submit the **.html** file and **.ipynb** notebook on Canvas.

Note: You will lose points if the notebook is not structured properly or if all the cells are not already run before converting to HTML.