Project 1

Jinting Zhang 910215-2668

**Question 1:** Show the correlation among the attributes in the dataset. Then, find two features with the highest correlation to the target, "Concrete compressive strength (Mpa)".

Answer:

dataset.corr()

**Question 2:** Separate the dataset into the output target which is "Concrete compressive strength (Mpa)" (y\_dataset) and the rest as the input features (X\_dataset).

Answer:

y\_dataset = dataset.pop("Concrete compressive strength (Mpa)")

X\_dataset = dataset

**Question 3:** Find the normalized input features.

Answer:

X\_dataset\_mean = X\_dataset.mean()

X\_dataset\_std = X\_dataset.std()

X\_dataset\_norm = (X\_dataset - X\_dataset\_mean)/X\_dataset\_std

**Question 4:** Split the dataset into the training set (80%) and validation/test set (20%). Make sure that the random state is set to 100 or random\_state=100 for reproducibility.

Answer:

X\_train\_norm, X\_test\_norm, y\_train, y\_test = train\_test\_split(X\_dataset\_norm, y\_dataset, test\_size=0.2,random\_state=100)

**Question 5:** Train the initialized model using the training data.

Answer:

model\_LR.fit(X\_train\_norm, y\_train)

**Question 6:** After training, evaluate the performance of the linear regression model against the test set. To do so, run the following cell and report the mean\_squared\_error.

Answer:

113.1787593778991

**Question 7:** Complete the code and define a neural network with two more hidden layers with 16 and 8 nodes. Set the activation function for the hidden layers as relu and the kernel\_regularizer as l2 with l=0.001.

Answer:

layers.Dense(16, activation='relu', kernel\_regularizer=tf.keras.regularizers.l2(l=0.001)),

layers.Dense(8, activation='relu', kernel\_regularizer=tf.keras.regularizers.l2(l=0.001)),

**Question 8:** Train the neural network model. Set epochs to 200 and batch\_size to 30.

Answer:

model\_NN.fit(X\_train\_norm, y\_train, epochs=200, batch\_size=30,validation\_split= 0.2)

**Question 9:** Evaluate the performance of the neural network model against the test set. To do so, run the following cell and report the mean\_squared\_error.

Answer:

41.6842

**Question 10:** Which model did perform better against the test set in predicting the output. Moreover, is there any overfitting observed for either the multivariate linear regression model or the neural network model? Motivate your answer.

Answer: The comparison between performance of the two models (i.e. the linear

regression and the neural network) is simply done based on the evaluations on

the test set. Here the neural network model 41.6842 performs better against the test set in predicting the output compared to the linear regression model 113.1787593778991. The overfitting issues because the MSE against the test set is just lower than the training set for the multivariate linear model regression.