

Results

Task1

- The dataset was divided into 2000 training data points and 300 testing data points.
- The neural network consists of 4 layers.
- GridSearchCV of scikit-learn was used to pick the best set of hyperparameters.
- The model is shown to be optimized for batch_size and optimizer function, but was in fact tested for epoch and loss functions manually and could not be shown as part of grid search as it took a lot of time for training different combinations on hyperparameters.
- The range of hyperparameters tested were
 - batch_size = [15,25, 30]
 - optimizer = ['adam', 'rmsprop', 'adagrad']
 - epoch= [600,800,900]
 - loss=['mse','mae','mean_squared_logarithmic_error']
- The evaluation metric is set of MSE to set a standard evaluation metric for both grid search as well as for the regression comparison.
- Once the best set of features were obtained from grid search, the optimal hyperparameters were used to train the optimal model.

```
Best: -0.000007 using {'optimizer': 'adam', 'batch_size': 15}
-0.000007 (0.000004) with: {'optimizer': 'adam', 'batch_size': 15}
-18.689168 (26.420360) with: {'optimizer': 'rmsprop', 'batch_size': 15}
-37.353989 (26.423076) with: {'optimizer': 'adagrad', 'batch_size': 15}
-18.689075 (26.420426) with: {'optimizer': 'adam', 'batch_size': 25}
-18.663179 (26.413146) with: {'optimizer': 'rmsprop', 'batch_size': 25}
-18.691090 (26.422984) with: {'optimizer': 'adagrad', 'batch_size': 25}
-18.663017 (26.413260) with: {'optimizer': 'adam', 'batch_size': 30}
-37.352084 (26.421786) with: {'optimizer': 'rmsprop', 'batch_size': 30}
-18.689235 (26.420313) with: {'optimizer': 'adagrad', 'batch_size': 30}
```

- The final model picked was
 - MSE = 0.000007 using {'optimizer': 'adam', 'batch_size': 15}
- The overall model chosen is -

Layer (type)	Output Shape	Param #
dense_13 (Dense)	(None, 64)	384
dense_14 (Dense)	(None, 32)	2080
dense_15 (Dense)	(None, 8)	264
dense_16 (Dense)	(None, 1)	9
Total params: 2,737		
Trainable params: 2,737		
Non-trainable params: 0		

- The MSE of testing data containing 300 data points was : **9.425**
- We see that the model performs well on testing data, but when compared to training data which is expected.
- We can also conclude that grid search is not an optimal method for hyperparameter selection as it takes a very long time to train.
- We stop at 800 epochs as we do not see a significant reduction in MSE, but very low loss values.

Task2

- We compare the neural network model with a multiple regression model.
- We fit the 2000 data points to an OLS estimator as done previously in project and we see very similar coefficients.

	coef
X1	9.0617
X2	3.4277
X3	3.3952
X4	6.2145
X5	7.0392

- We obtain an testing data MSE = **223.681**, which is very high when compared to MSE of the neural network model.
- Comparing the MSE of the 2 models, we can say that the chosen optimal neural network model is a better suited model for this data.