CS 6375
ASSIGNMENT 1 Report

Names of students in your group:
Yash Nadkarni (ydn200000)
Soham Savalapurkar (sxs200389)

Number of free late days used: 0

Details about our Dataset:
For assignment 1, we used the Airfoil Self-Noise dataset from the UCI website. This dataset has 5 independent variables and 1 dependent (target) variable. It has 1503 instances.
Independent Variables:
1. Frequency
2. Angle of Attack
3. Chord length
4. Free stream velocity
5. Suction side displacement thickness
Target Variable:
1. Scaled sound pressure level

Log of trials:

Sr. no	Epochs	Learning Rate	Training MSE	Testing MSE	Testing R^2
1	100	0.002	10526.23	10478.96	-247.8
2	500	0.002	2143.54	2127.13	-49.51
3	1000	0.002	311.03	304.22	-6.22
4	1000	0.003	62.84	58.55	-0.39
5	1000	0.004	29.12	25.76	0.38
6	1000	0.005	24.49	21.45	0.49
7	1000	0.006	23.83	20.9	0.503
8	1500	0.006	23.65	20.8	0.505
9	2000	0.006	23.63	20.78	0.506
10	3000	0.006	23.62	20.77	0.506

Answers to the questions in assignment:

1. Are you satisfied that you have found the best solution? Explain.

Answer: Yes. Considering the fact that the r2 score of the model used with the help of the scikit learn package is very close to the r2 score of the model we created, it can be concluded that we have indeed found the best solution.

2. Are you satisfied that the package has found the best solution? How can you check? Explain.

Answer: No. I would say that the package has given satisfactory results compared to our model. However, it is important to mention that a r2 score should be as close to 1 as possible. In that regard, there is some room for improvement in the model.
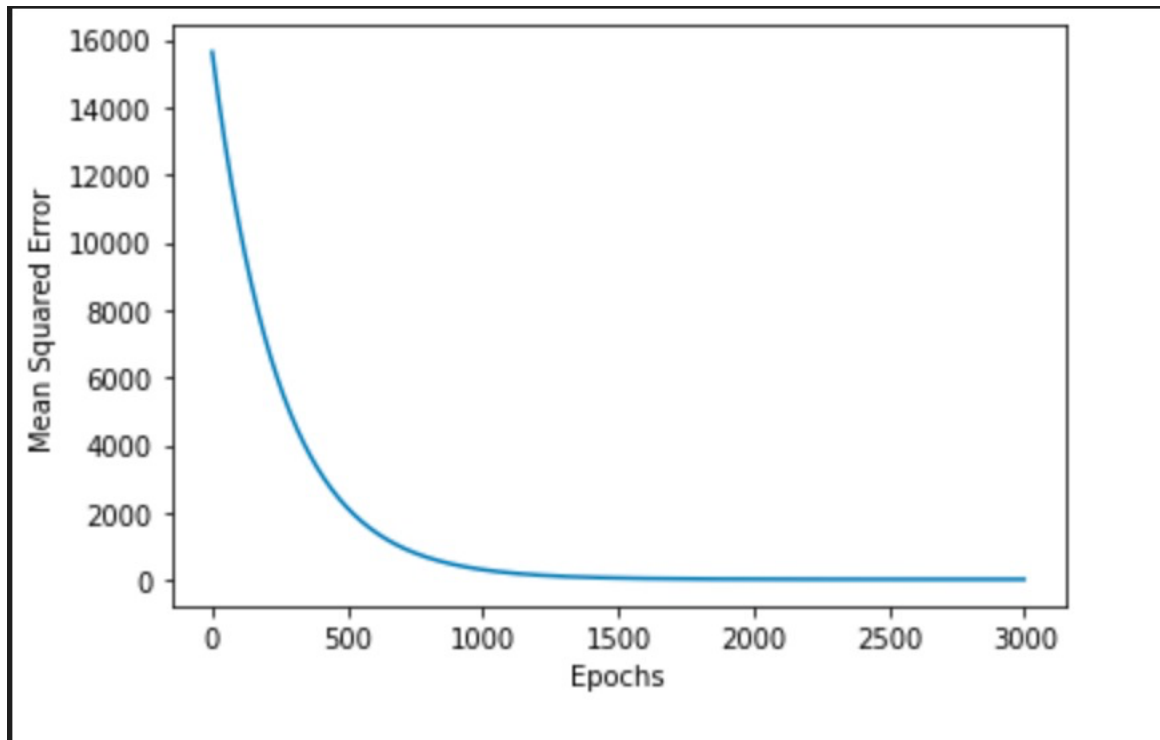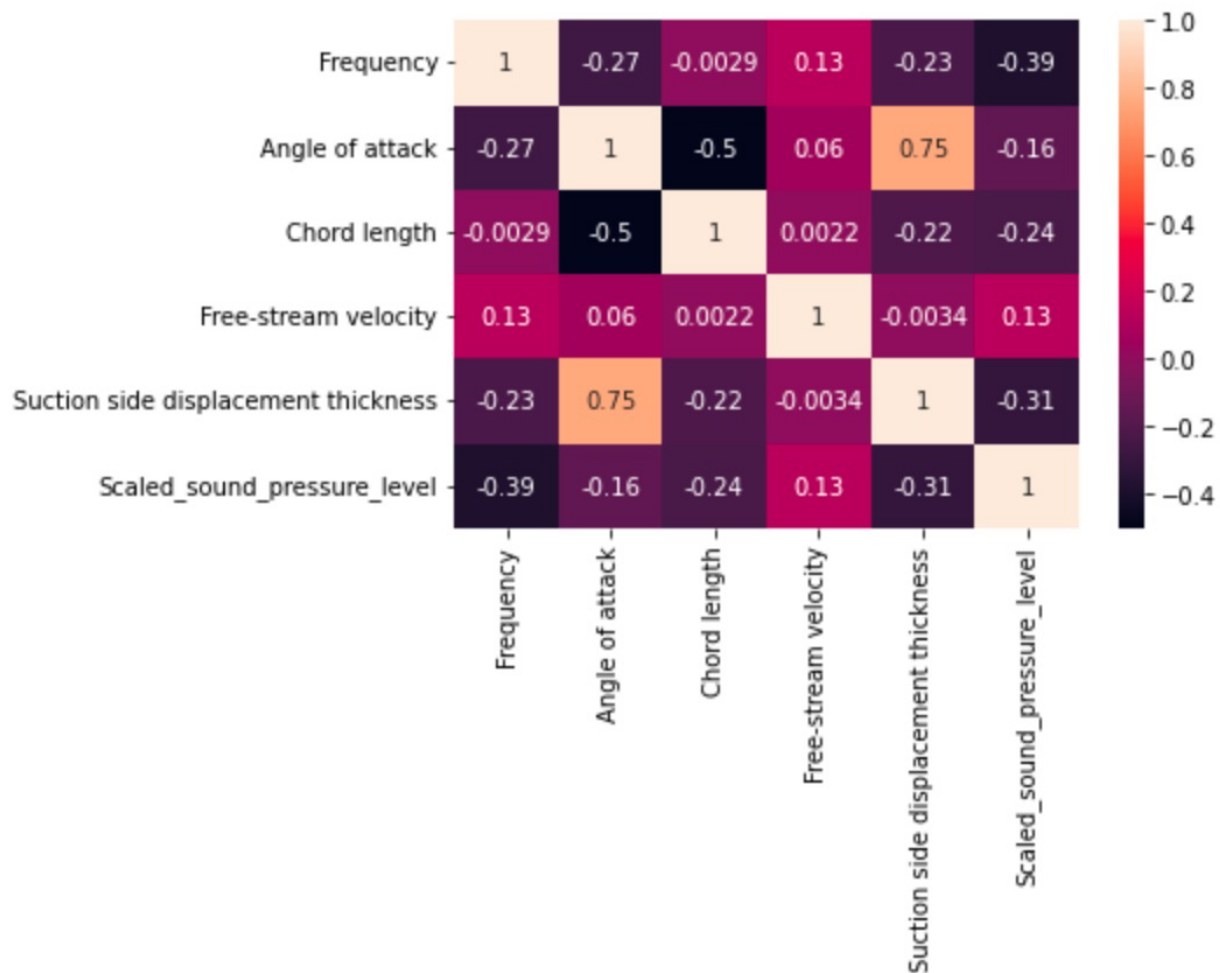
Plots:

1. R2 Score vs Epochs:
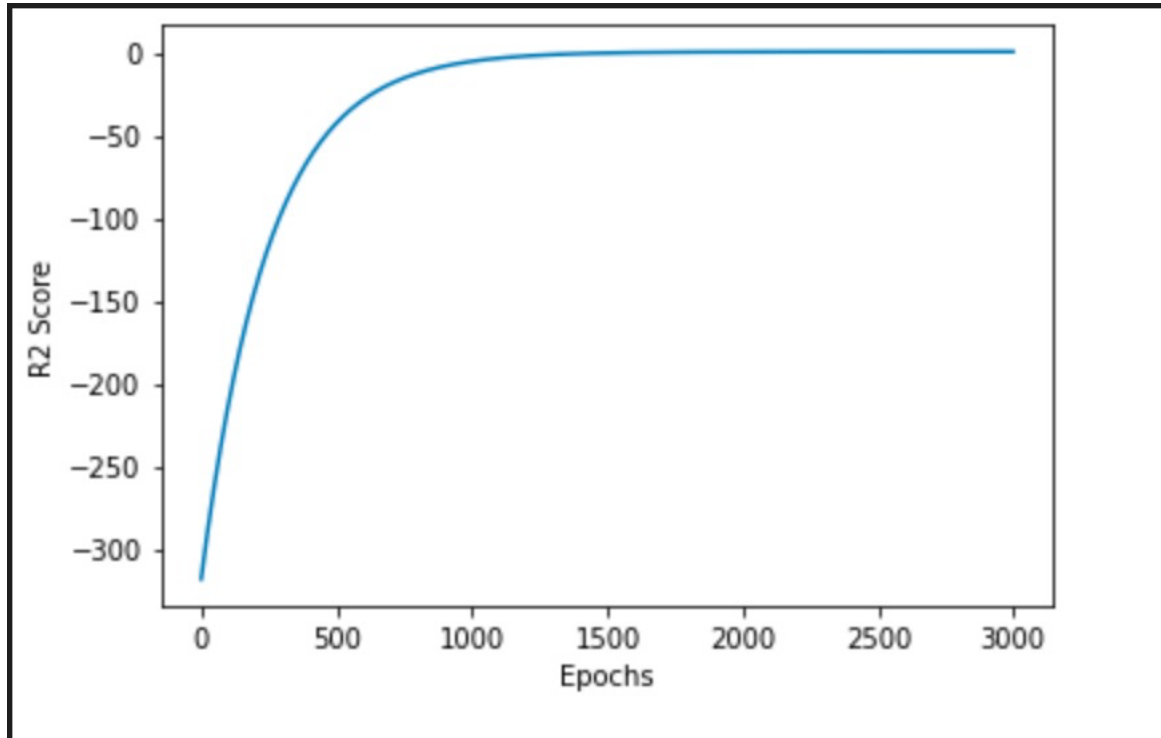
As we can see from the above diagram, the r2 score barely decreases as we go above 1500 epochs.

2. Mean Squared Error vs Epochs:

Similarly, since the mean square error fractionally decreases after 1500 epochs, we chose 1500 epochs for our model training.

3. Heatmap

From this heatmap, we can clearly get an idea of the correlation between the independent variables and the target variables. Since no feature appears to have a stronger correlation as compared to others w.r.t. the target variable, we decided to use all the features.