Project 1

ET1550 Introduction to Machine Learning and Artificial Intelligence

Kevin Rasmusson

Lund, 07/10/2022

Question 1

Code

```
# Built in function to show statistical information about dataset
dataset.describe()
```

Output

	Plant_Air_NM3	Instrument_Air_NM3	Steam_Ton	Power_MWh	Production_N2_NM3	Production_O2_NM3	Production_CA_NM3	Total_Production_NM3	Energy_Input_MJ	EnPI_MJ_NM3
count	35.00	35.00	35.00	35.00	35.00	35.00	35.00	35.00	35.00	35.00
mean	4,900,789.91	12,638,141.43	10,806.14	110,771.86	37,896,411.66	85,677,196.92	17,551,953.34	141,125,561.92	434,035,286.62	3.10
std	957,913.41	1,010,165.44	395.09	6,774.74	2,952,302.19	10,204,705.41	1,754,731.93	11,010,328.79	24,960,544.57	0.15
min	3,681,978.00	11,183,877.00	9,734.00	90,951.00	32,168,031.00	55,952,502.00	15,157,719.00	104,933,052.00	360,892,509.70	2.83
25%	4,091,106.00	11,818,519.00	10,560.50	108,650.50	35,796,031.00	79,522,018.50	16,310,828.00	135,511,752.00	427,025,746.80	3.02
50%	4,638,478.00	12,309,519.00	10,811.00	111,556.00	38,239,531.00	88,990,502.00	16,804,019.00	145,100,367.00	437,452,996.00	3.07
75%	5,416,728.00	13,618,519.00	11,096.00	116,016.00	39,836,531.00	93,915,002.00	19,149,019.00	149,192,052.00	453,791,040.80	3.19
max	6,844,478.00	14,301,519.00	11,624.00	118,584.00	45,881,531.00	99,535,502.00	21,159,019.00	155,448,952.00	462,540,327.50	3.46

Question 2

Code

```
dataset.corr()
```

Output

	Plant_Air_NM3	Instrument_Air_NM3	Steam_Ton	Power_MWh	Production_N2_NM3	Production_O2_NM3	Production_CA_NM3	Total_Production_NM3	Energy_Input_MJ	EnPI_MJ_NM3
Plant_Air_NM3	1.00	0.59	-0.01	0.42	-0.28	0.37	0.89	0.41	0.41	-0.21
Instrument_Air_NM3	0.59	1.00	-0.35	0.34	-0.43	0.49	0.90	0.48	0.32	-0.42
Steam_Ton	-0.01	-0.35	1.00	0.42	0.52	0.13	-0.20	0.23	0.46	0.16
Power_MWh	0.42	0.34	0.42	1.00	0.06	0.80	0.43	0.82	1.00	-0.20
Production_N2_NM3	-0.28	-0.43	0.52	0.06	1.00	-0.13	-0.40	0.08	0.08	-0.06
Production_O2_NM3	0.37	0.49	0.13	0.80	-0.13	1.00	0.49	0.97	0.79	-0.69
Production_CA_NM3	0.89	0.90	-0.20	0.43	-0.40	0.49	1.00	0.50	0.41	-0.36
Total_Production_NM3	0.41	0.48	0.23	0.82	0.08	0.97	0.50	1.00	0.82	-0.71
Energy_Input_MJ	0.41	0.32	0.46	1.00	0.08	0.79	0.41	0.82	1.00	-0.19
EnPI_MJ_NM3	-0.21	-0.42	0.16	-0.20	-0.06	-0.69	-0.36	-0.71	-0.19	1.00

Question 3

Code

```
dataset_mean = dataset.mean()
dataset_std = dataset.std()
dataset_norm = (dataset - dataset_mean)/dataset_std
```

Output

Dataset normalized.

Question 4

Code

```
train_set_norm, test_set_norm = train_test_split(dataset_norm, train_size=0.8,
test_size=0.2, random_state=100)
```

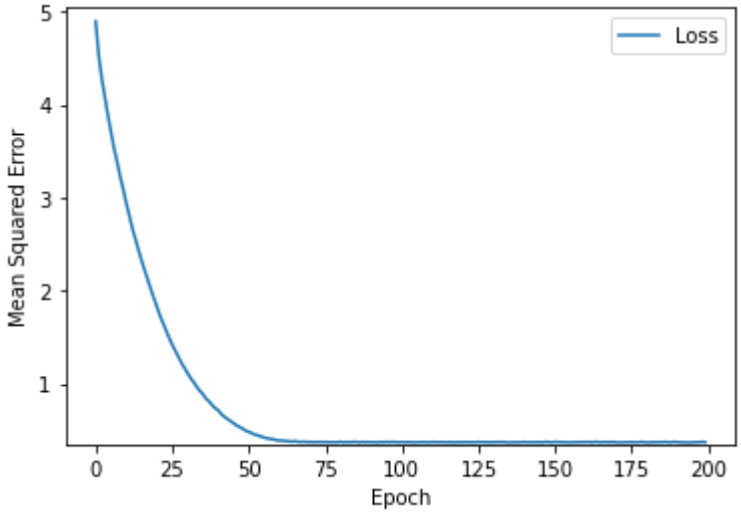
Output
Dataset split.

Question 5

Code

```
learning_rate = 0.01 # given as adequate in hint
epochs = 200 # hint was “a couple of hundreds”
batch_size = 7 # just because divisible by 0.8*35=28
```

Output



The learned weight for your model is [[0.16677113]
[0.8308066]]
The learned bias for your model is [-0.06133229]

Question 6

A - What are the trained model’s parameters?
The parameters of the model is given in the output in question 5, i.e., the reported weights and biases. The weights are given in an array, where element at index 0 is w_1 and element at index 1 is w_2. Therefore, the parameters are as follows: w_0=*-0.06133229,* w_1=*0.16677113* and w_2=*0.8308066*

B - Report the model’s performance in terms of the mean squared error.
1/1 [==========] - 0s 121ms/step - loss: 0.3881 - mean_squared_error: 0.3881
[0.3880944848060608, 0.3880944848060608]
Out[19]: [0.3880944848060608, 0.3880944848060608]
(To clarify, the mean squared error is ~0.3881)

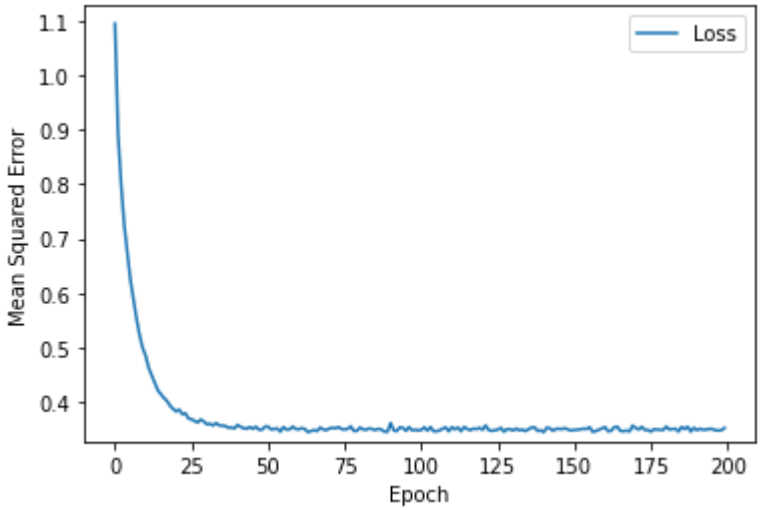
Question 7

Code

```
instrument_air = tf.feature_column.numeric_column("Instrument_Air_NM3")
```

```
feature_columns.append(instrument_air)
```

New outputs with the new feature



The learned weight for your model is [[-0.18285407]
 [0.25368664]
 [0.89165896]]
The learned bias for your model is [-0.05476298]

Evaluate the trained linear regression model against the test set:
/1 [==============================] - 0s 161ms/step - loss: 0.3468 - mean_squared_error: 0.3468
[0.34683582186698914, 0.34683582186698914]

Comment

When comparing the two models we realize that the second one (using three features) results in a lower mean squared error, implying that the second model is an improvement when compared to the first one. As we know, the MSE for model1 was 0.3881, which is greater than model2's MSE of 0.3468.