HW6 Report
Gina Roh

1. Linear regression: Fit a line to go very near the 2017-2021 citation columns, minimizing MSE. Use that line to predict the 2022 citation numbers, as in HW4.

The mean square error for the linear regression (LR) model is 22318.72. Surprisingly, despite its simplicity, the LR model consistently delivers strong performance. The mean square error for the neural network (NN) model tends to vary depending on the experiment. While the LR model's performance remains stable, the performance of the NN model fluctuates. In some cases, the LR model outperforms the NN model, but there are instances where it does not.

Experiment 1.

Task1. Regression

	univ_rank	first_initial	last_initial	cit_2017	cit_2018	cit_2019	cit_2020	cit_2021	cit_2022	h_index	i_10_index	predict_LR	predict_NN
0	51	I	P	38	102	159	245	277	381	16	24	331.956383	337.488342
1	51	S	M	153	333	510	749	963	1048	31	59	1110.242402	1105.813965
2	51	A	B	5524	8950	12526	14204	16734	17508	60	88	17495.222797	19010.160156
3	51	W	H	161	183	206	215	179	262	22	33	180.198893	209.633652
4	51	F	N	70	96	88	133	157	156	16	24	193.009931	163.180161
5	51	M	I	238	386	641	602	1025	1249	41	110	1131.572974	1061.630005
6	51	R	F	41	115	210	312	473	554	15	21	570.777614	527.289795
7	51	S	J	54	72	113	139	144	141	9	9	172.272102	179.815765
8	51	S	Z	135	92	160	184	238	332	24	34	280.677799	275.249146
9	51	J	Z	1678	2066	2635	3253	4319	4125	23	30	4726.929948	4617.534668
10	52	E	G	151	147	156	152	169	167	31	85	183.088679	170.515961
11	52	M	C	85	121	202	264	376	383	28	65	445.295386	420.399231
12	52	W	E	1375	1264	1038	998	947	784	37	65	835.326749	808.786865
13	52	A	D	183	286	356	395	449	490	22	31	479.024626	497.092560
14	52	R	C	89	128	103	109	108	103	20	39	119.311579	97.284363
15	52	V	C	19	22	52	116	172	188	13	21	231.936856	199.082474
16	52	T	B	503	463	584	722	945	893	46	163	1051.006748	1010.132568
17	52	W	A	47	82	98	128	178	346	17	35	217.976548	187.076019
18	52	K	A	139	125	84	80	74	47	16	24	80.991960	49.791828
19	52	S	H	205	201	220	210	202	187	21	29	203.302863	215.106812

MSE_LR: 22318.72
MSE_NN: 129482.96

Experiment 2.

Task1. Regression

	univ_rank	first_initial	last_initial	cit_2017	cit_2018	cit_2019	cit_2020	cit_2021	cit_2022	h_index	i_10_index	predict_LR	predict_NN
0	51	I	P	38	102	159	245	277	381	16	24	331.956383	314.540802
1	51	S	M	153	333	510	749	963	1048	31	59	1110.242402	1082.714844
2	51	A	B	5524	8950	12526	14204	16734	17508	60	88	17495.222797	17569.634766
3	51	W	H	161	183	206	215	179	262	22	33	180.198893	166.300919
4	51	F	N	70	96	88	133	157	156	16	24	193.009931	180.043106
5	51	M	I	238	386	641	602	1025	1249	41	110	1131.572974	1073.422119
6	51	R	F	41	115	210	312	473	554	15	21	570.777614	536.638062
7	51	S	J	54	72	113	139	144	141	9	9	172.272102	148.246796
8	51	S	Z	135	92	160	184	238	332	24	34	280.677799	226.113480
9	51	J	Z	1678	2066	2635	3253	4319	4125	23	30	4726.929948	4597.134766
10	52	E	G	151	147	156	152	169	167	31	85	183.088679	159.620163
11	52	M	C	85	121	202	264	376	383	28	65	445.295386	407.719086
12	52	W	E	1375	1264	1038	998	947	784	37	65	835.326749	839.771484
13	52	A	D	183	286	356	395	449	490	22	31	479.024626	473.254059
14	52	R	C	89	128	103	109	108	103	20	39	119.311579	118.824295
15	52	V	C	19	22	52	116	172	188	13	21	231.936856	198.773270
16	52	T	B	503	463	584	722	945	893	46	163	1051.006748	966.756531
17	52	W	A	47	82	98	128	178	346	17	35	217.976548	201.282394
18	52	K	A	139	125	84	80	74	47	16	24	80.991960	67.102905
19	52	S	H	205	201	220	210	202	187	21	29	203.302863	180.241791

MSE_LR: 22318.72
MSE_NN: 15785.89

2. Logistic regression: Classify individuals into 3 categories, as in HW5.

The logistic regression model achieves a remarkable 100% accuracy when classifying the test set. This is impressive, considering its simplicity. It outperforms the neural network model from HW5, which achieves 0.8 accuracy in this case. Similar to the previous task, the accuracy of the neural network varies depending on the experiments.

Experiment 1.

```
Task2. Classification
   univ_rank first_initial last_initial  cit_2017  cit_2018  cit_2019  cit_2020  cit_2021  cit_2022  h_index  i_10_index  ratio  category  predict_LR  predict_NN
0         51             I            P        38       102       159       245       277       381       16          24   1.38         2           2           2
1         51             S            M       153       333       510       749       963      1048       31          59   1.09         1           1           0
2         51             A            B      5524      8950     12526     14204     16734     17508       60          88   1.05         1           1           0
3         51             W            H       161       183       206       215       179       262       22          33   1.46         2           2           2
4         51             F            N        70        96        88       133       157       156       16          24   0.99         0           0           0
5         51             M            I       238       386       641       602      1025      1249       41         110   1.22         2           2           2
6         51             R            F        41       115       210       312       473       554       15          21   1.17         2           2           2
7         51             S            J        54        72       113       139       144       141        9           9   0.98         0           0           0
8         51             S            Z       135        92       160       184       238       332       24          34   1.39         2           2           2
9         51             J            Z      1678      2066      2635      3253      4319      4125       23          30   0.96         0           0           0
10        52             E            G       151       147       156       152       169       167       31          85   0.99         0           0           0
11        52             M            C        85       121       202       264       376       383       28          65   1.02         0           0           0
12        52             W            E      1375      1264      1038       998       947       784       37          65   0.83         0           0           0
13        52             A            D       183       286       356       395       449       490       22          31   1.09         1           1           0
14        52             R            C        89       128       103       109       108       103       20          39   0.95         0           0           0
15        52             V            C        19        22        52       116       172       188       13          21   1.09         1           1           2
16        52             T            B       503       463       584       722       945       893       46         163   0.94         0           0           2
17        52             W            A        47        82        98       128       178       346       17          35   1.94         2           2           2
18        52             K            A       139       125        84        80        74        47       16          24   0.64         0           0           0
19        52             S            H       205       201       220       210       202       187       21          29   0.93         0           0           0
Accuracy_LR: 1.00
Accuracy_NN: 0.80
```

Experiment 2.

```
Task2. Classification
   univ_rank first_initial last_initial  cit_2017  cit_2018  cit_2019  cit_2020  cit_2021  cit_2022  h_index  i_10_index  ratio  category  predict_LR  predict_NN
0         51             I            P        38       102       159       245       277       381       16          24   1.38         2           2           2
1         51             S            M       153       333       510       749       963      1048       31          59   1.09         1           1           2
2         51             A            B      5524      8950     12526     14204     16734     17508       60          88   1.05         1           1           0
3         51             W            H       161       183       206       215       179       262       22          33   1.46         2           2           0
4         51             F            N        70        96        88       133       157       156       16          24   0.99         0           0           0
5         51             M            I       238       386       641       602      1025      1249       41         110   1.22         2           2           2
6         51             R            F        41       115       210       312       473       554       15          21   1.17         2           2           2
7         51             S            J        54        72       113       139       144       141        9           9   0.98         0           0           0
8         51             S            Z       135        92       160       184       238       332       24          34   1.39         2           2           0
9         51             J            Z      1678      2066      2635      3253      4319      4125       23          30   0.96         0           0           0
10        52             E            G       151       147       156       152       169       167       31          85   0.99         0           0           0
11        52             M            C        85       121       202       264       376       383       28          65   1.02         0           0           2
12        52             W            E      1375      1264      1038       998       947       784       37          65   0.83         0           0           0
13        52             A            D       183       286       356       395       449       490       22          31   1.09         1           1           0
14        52             R            C        89       128       103       109       108       103       20          39   0.95         0           0           0
15        52             V            C        19        22        52       116       172       188       13          21   1.09         1           1           2
16        52             T            B       503       463       584       722       945       893       46         163   0.94         0           0           0
17        52             W            A        47        82        98       128       178       346       17          35   1.94         2           2           2
18        52             K            A       139       125        84        80        74        47       16          24   0.64         0           0           0
19        52             S            H       205       201       220       210       202       187       21          29   0.93         0           0           0
Accuracy_LR: 1.00
Accuracy_NN: 0.65
```