11/06/22
Sehyun Cho
CAP4612

Homework 4

- **Describe the algorithms/approaches/tools used:**
  - **a. <u>What it is or What it does</u>**
    Tools used: LinearRegression, Lasso for creating models, PolynomialFeatures for create models that changing N values, pipeline to apply each LinearRegression and Lasso model, and pyplot to plot models.
  - **b. <u>How it does & Application</u>**
    1. Import data x,y into two numpy arrays and create as a data frame.
       Figure 1 – Reading data

	x	y
0	-6.860121	0.8116
1	-4.324130	0.9072
2	-4.358625	0.9052
3	-4.358427	0.9039
4	-6.955852	0.8053
...	...	...
77	-5.132415	0.8964
78	-4.811353	0.8963
79	-4.098269	0.9074
80	-3.661743	0.9119
81	-3.264401	0.9228

82 rows × 2 columns

    2. Create plot graph for the initial dataset.

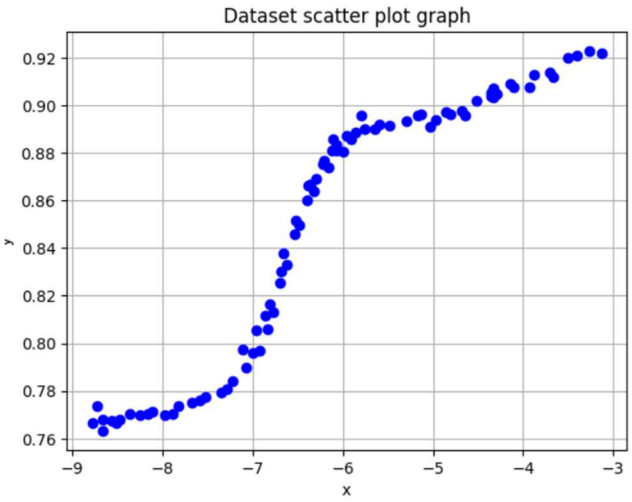

Figure 2 – Initial dataset scatter plot

3. Concatenate each terms until x10, to illustrate higher order term dataset.

Figure 3 – Create overfitted model dataset

	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	y
0	-6.860121	47.061259	-322.845927	2214.762094	-15193.535763	104229.492448	-715026.920996	4.905171e+06	-3.365007e+07	2.308435e+08	0.8116
1	-4.324130	18.698101	-80.853019	349.618968	-1511.797883	6537.210647	-28267.748970	1.222334e+05	-5.285532e+05	2.285533e+06	0.9072
2	-4.358625	18.997612	-82.803469	360.909276	-1573.068212	6856.414522	-29884.540122	1.302555e+05	-5.677349e+05	2.474544e+06	0.9052
3	-4.358427	18.995884	-82.792168	360.843598	-1572.710389	6854.543023	-29875.023648	1.302081e+05	-5.675025e+05	2.473418e+06	0.9039
4	-6.955852	48.383882	-336.551143	2341.000068	-16283.650894	113266.671807	-787866.248553	5.480281e+06	-3.812003e+07	2.651573e+08	0.8053
...	...	...	...	...	...	...	...	...	...	...	...
77	-5.132415	26.341680	-135.196427	693.884125	-3561.301065	18278.073839	-93810.654364	4.814752e+05	-2.471130e+06	1.268287e+07	0.8964
78	-4.811353	23.149115	-111.378556	535.881518	-2578.314991	12405.182802	-59685.709816	2.871690e+05	-1.381671e+06	6.647708e+06	0.8963
79	-4.098269	16.795811	-68.833758	282.099278	-1156.118813	4738.086246	-19417.953440	7.958000e+04	-3.261403e+05	1.336611e+06	0.9074
80	-3.661743	13.408360	-49.097966	179.784121	-658.323205	2410.610236	-8827.034604	3.232233e+04	-1.183561e+05	4.333894e+05	0.9119
81	-3.264401	10.656315	-34.786485	113.557040	-370.695725	1210.099533	-3950.250245	1.289520e+04	-4.209511e+04	1.374153e+05	0.9228

82 rows × 11 columns

4. Apply Linear Regression for each model, until N=10

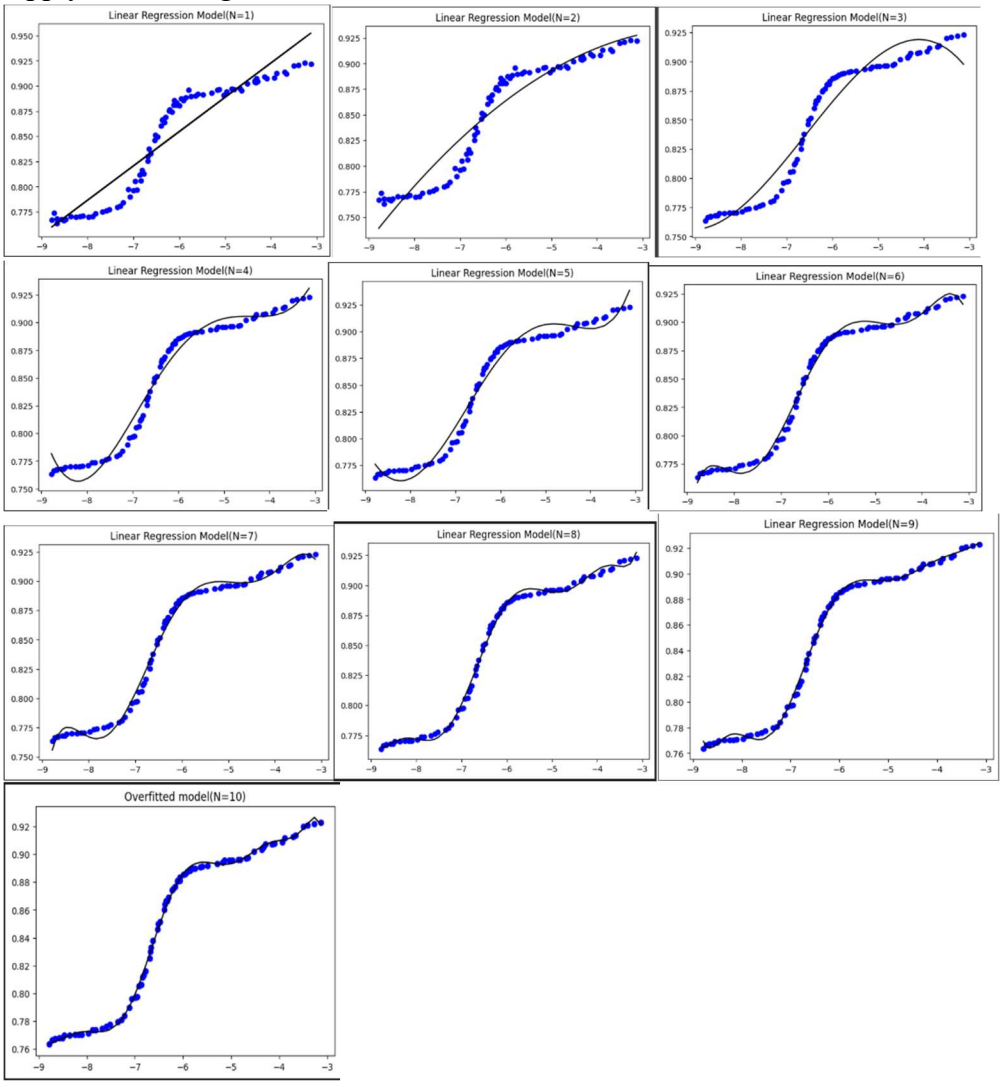

Figure 4 – Create Linear Regression models

5. Apply LASSO to reduce terms from the overfitted model, acquire models with changing alpha value.

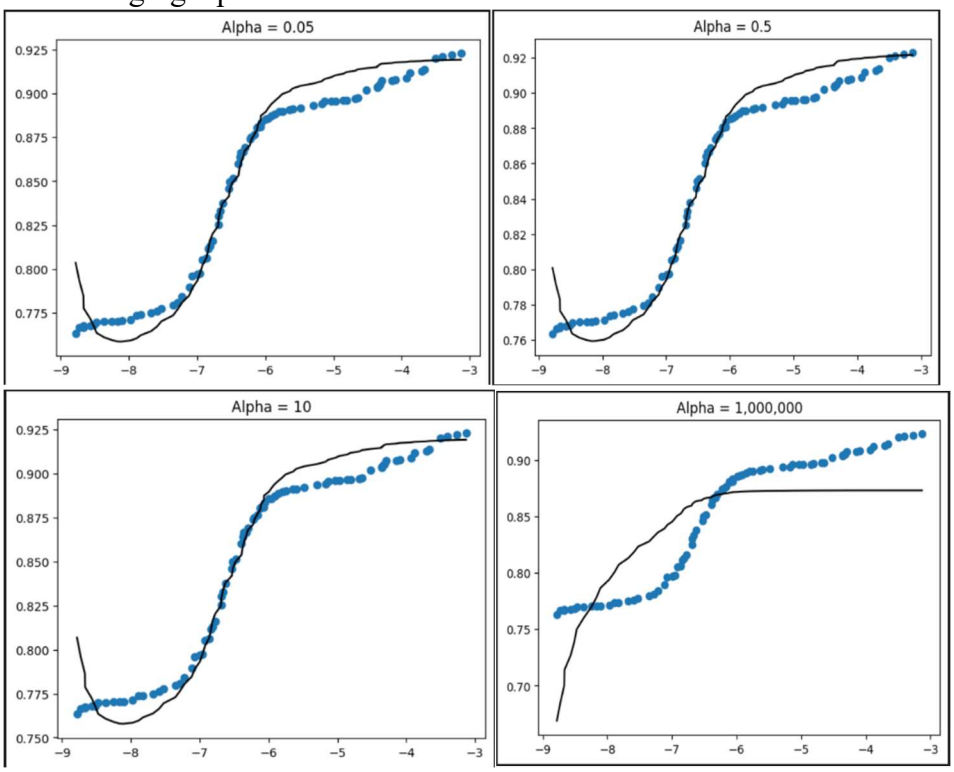

Figure 5 – Models applying LASSO with different alpha parameter values

- **Describe results:**

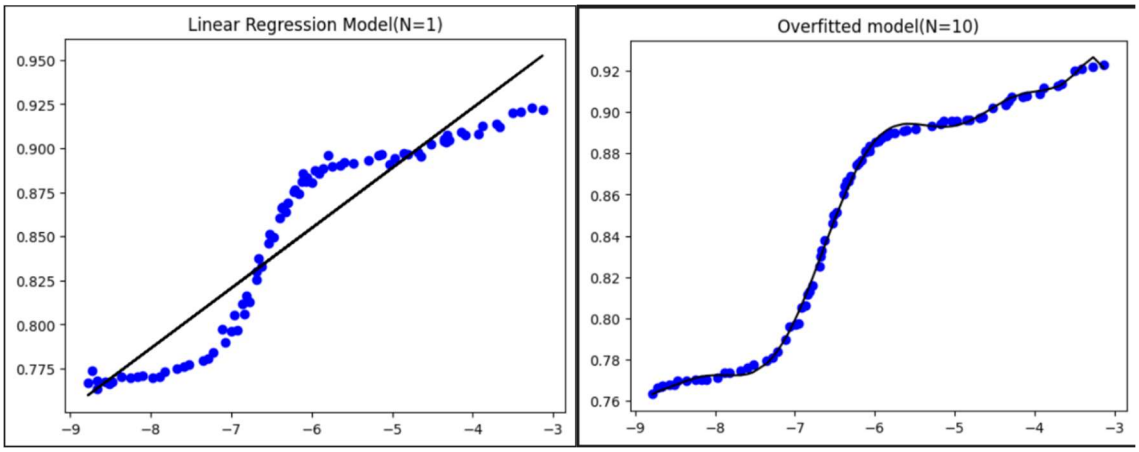

Figure 4 – Create Linear Regression models

1. **Describe the figure and table.**
   Two Linear regression models when N=1 and N=10.
2. **Your observation about the figure and table.**
   I could observe when order of terms getting higher, the model getting overfitted.
3. **Conclusion.**
   In conclusion, testing Linear Regression by creating graphs with different terms successfully done.

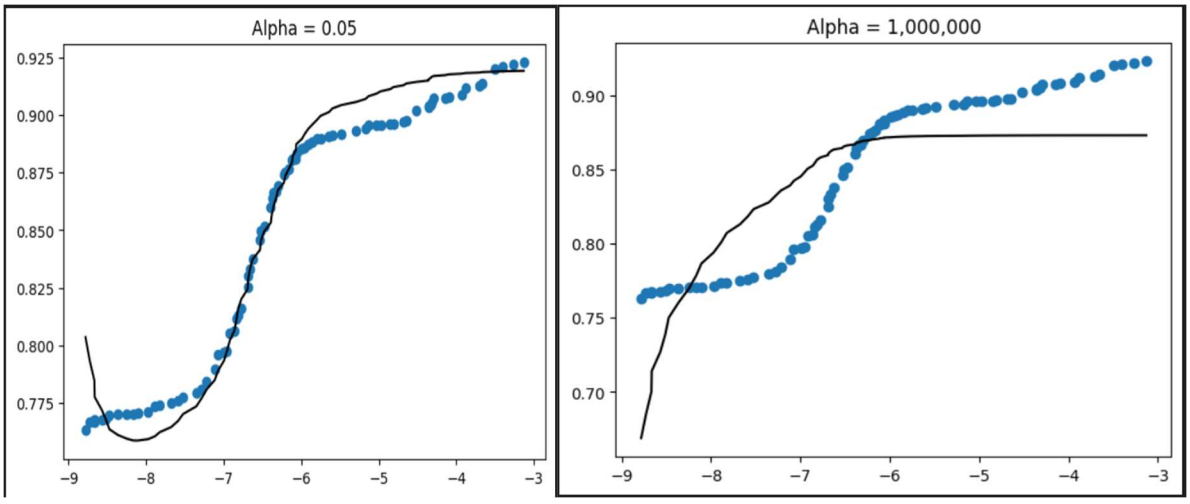

Figure 5 – Models applying LASSO with different alpha parameter values

1. **Describe the figure and table.**
   Two LASSO models when alpha is 0.05 and 1,000,000.
2. **Your observation about the figure and table.**
   I could observe size of the alpha reduce the model complexity.
3. **Conclusion.**
   In conclusion, LASSO models are reducing complexity of overfitted model, and above plots shows when setting higher alpha value, will obtain underfitted model.