

ASSIGNMENT-2 REPORT

As multiple local minima are possible in polynomial regression, multiple optimal solutions to minimizing the cost function are observed based on the initialization of the weights.

All the data was initially normalized and new features were generated according to the degree of the polynomial to be fit. An N-th degree polynomial has $(N+1)C2$ coefficients.

Learning rate: (1e-6)

Stopping criteria: ($E-E' \leq 5e-2$)

Maximum iterations: 50000

Gradient Descent:

DEG	TRAIN ERR	AVG TEST ERR	R2 ERROR	RMSE	WEIGHTS(np.random.randn())
1	2504.969060	0.0082859983	2.5010214167	0.1287322	[0.20848472 0.09536773 -0.09982884]
2	2416.236659	0.008031471	6.212487647	0.1267396	[0.1423993 0.52602894 -0.47462574 -0.17928082 -0.03033481 0.13114884]
3	2243.467785	0.007348374	12.65263576	0.1212301	[0.22956625 -0.13999825 1.79395354 -2.04201335 -0.53483783 -1.52421099 1.46343568 1.93033192 0.42746262 -1.5781994]
4	2165.918867	0.007224321	15.68624651	0.1202025	[0.20091943 -0.00290968 1.27206105 -0.40164413 -1.72933962 -0.21291626 -1.98774206 1.08462712 1.66735796 0.57902535 0.02645383 0.27884248 1.56978226 -0.93956409 -1.40090874]
5	2124.145589	0.00695314	17.6060985	0.1179249	[0.17442318 0.00980305 1.31990218 -0.28355097 -1.4693503 -0.94444191 0.08125487 -1.98970689 0.69908159 1.32451212 1.63676402 -0.2194968 -0.29283898 0.28029615 0.39492675 1.56601267 -0.10389244 -0.71290387 0.67861609 -1.26979099 -0.87991535]
6	2110.53817	0.00695059	17.77905469	0.1179032	[0.17048744 -0.08962674 1.50332933 -0.02050638 -1.53132001 -1.52887083 -0.15819974 0.19751607 -1.95866819 0.57652114 0.99373625 1.036232 1.56734689 -0.37704709 -0.29113032 0.29275535 0.31045194 0.60387746 1.19626403 0.06177757 -0.36096084 0.31454571 0.73898155 -0.59345753 -0.86432248 -0.03954899 -0.88730104 -0.15256575]

From the above data:

- We can observe that the model has converged to some local minima.
- As higher degree polynomials are fit to the data, the training error decreases while the R2 error increases, indicating some overfitting.
- Over fit: degree 6 polynomial
- Best fit: degree 4 polynomial

With L1 Regularization:

<u>DEG</u>	<u>VAL ERR</u>	<u>AVG TEST ERR</u>	<u>R2 ERROR</u>	<u>RMSE</u>	<u>REG. COEFF.</u>	<u>WEIGHTS</u>
6	645.36448	0.007294469	14.2374103	0.12078467	0.875	[0.18457507 -0.07398244 1.71922973 -1.7966707 0.44542585 -0.806041 -0.581372 -0.11608601 -0.81777302 -0.09801393 -0.87337241 0.40464632 2.5543963 -0.12994462 1.07814396 0.33470576 -0.42566046 -0.09357612 0.6889279 -0.36274973 0.72563591 -0.44936528 0.10844836 -0.27020638 0.10681549 -0.79221959 -0.63772207 0.25556698]

After L1-regularization:

- Regularization coeff = 0.875. The accuracy after regularization is close to that of degree 4 polynomial.

With L2 Regularization:

<u>DEG</u>	<u>VAL ERR</u>	<u>AVG TEST ERR</u>	<u>R2 ERROR</u>	<u>RMSE</u>	<u>REG. COEFF.</u>	<u>WEIGHTS</u>
6	648.47769	0.007259853	11.9812831	0.12049774	0.0	[0.16959764 0.18025695 0.65518339 -0.06978974 -0.58611338 -0.57898445 -0.16013752 -0.31185269 -0.4559507 -0.43000127 -0.33544481 0.07765144 0.75703779 0.55313306 -0.15820445 -0.19200157 0.24284709 1.00197346 0.63461067 -0.08260375 0.11207165 0.82501614 0.15354349 -0.17163259 0.36499707 -0.42624223 -0.25293941 -0.89574433]

After L2-regularization:

- Regularization coeff = 0, indicating no regularization in this case will yield a better answer.

With L2 Regularization:

<u>DEG</u>	<u>TRAIN ERR</u>	<u>AVG TEST ERR</u>	<u>R2 ERROR</u>	<u>RMSE</u>	<u>REG. COEFF.</u>	<u>WEIGHTS</u>
1						
2						
3						
4						
5						
6						