FOUNDATIONS OF DATA SCIENCE – A3 REPORT

Course ID: CS F320

Group members:

Parth Krishna Sharma: 2017B3A70907H

Siddhi Burse: 2017B3A70972H

Maurya Grover: 2017B3A71433H

NOVEMBER 5, 2020

Table of Contents

I. Overview

One of the important aspects of training a machine learning model is to avoid overfitting. As the number of training data points increase, the training error of the model keeps increasing. But it is important to keep in mind that the goal is to build a model that captures the true relation between the dependent and independent variables and not simply fit the training data.

To take into account this possibility of overfitting, we introduce regularization in our regression models. We will be using 2 types of regularization techniques:

1. Lasso Regression (least absolute shrinkage) aka L1 regression
2. Ridge Regression aka L2 regression

The report contains a comparative analysis of the models built.
Programming language used: Python
Libraries used: NumPy, Pandas, Matplotlib, mpl_toolkits and sklearn

II. Description of Data

The data used for building the models consists of 1338 datapoints. Each datapoint has 3features and 1target. The target variable is the insurance amount and the feature vectors for predicting the target variable are age and bmi.

The given features of the data lie in different ranges, for example, the range for age is 18-64, the range of bmi is 15-53 and insurance amount varies from Rs.1122 to Rs.63770. Hence, we bring the dataset to a common scale without distorting the differences in the values by normalizing it. We normalize the data by dividing by the maximum value of each feature. Therefore, on normalization, each value is brought into the range 0-1.

The data is shuffled and split into training data (70%), validation data (20%) and testing data (10%).

age	bmi	charges
0.968750	0.602579	0.730278
0.921875	0.517598	0.195451
0.437500	0.447958	0.061471
0.281250	0.737060	0.026090
0.937500	0.345097	0.210955
0.703125	0.540184	0.128257
0.859375	0.709863	0.480304
0.546875	0.491718	0.083524
0.296875	0.480990	0.026082
0.281250	0.514963	0.274451

III. Model description

We first build models using gradient descent and stochastic gradient descent methods for degrees of polynomial 1 to 10.

The hyperparameter values for gradient descent:
learning_rate = 0.01
epochs = 1000
precision = 0.000001

The hyperparameter values for stochastic gradient descent:
Learning_rate = 0.0005
Epochs = 100
Precision = 0.000001

We then introduce a regularization term and perform lasso and ridge regression on the data for degrees of polynomial from 1 to 10.

The learning rate, epochs and precision values are kept same as above and the regression is run for 10 randomly generated values of lambda (regularization constant).

Lasso regularization adds the **absolute magnitude** of weights as penalty term to the cost function.

$$\sum_{i=1}^{n}(Y_i - \sum_{j=1}^{p} X_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

Cost function

On the other hand, ridge regression adds **squared magnitude** of weights as penalty term to the cost function.

$$\sum_{i=1}^{n}(y_i - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} \beta_j^2$$

Cost function

The key difference between these two techniques is that Lasso regularization shrinks the less important feature's coefficient to zero. Hence, removing some features altogether. Therefore, this regularization technique works better for feature selection when there are a large number of features.

IV. Varying degree of polynomials

Training and testing RMSE (root mean square error) for regression with degree of polynomial varying from1 to 10 using 2 methods- Gradient descent and Stochastic Gradient descent:

	Gradient_descent		Stochastic_Gradient_descent	
Degree of Polynomial	**Train_rmse**	**Test_rmse**	**Train_rmse**	**Test_rmse**
1	0.187897554	0.23268771	0.277216911	0.3344456
2	0.183397659	0.2075169	0.268903744	0.303600101
3	0.183122257	0.20712183	0.265641118	0.300336717
4	0.182963592	0.2074689	0.263330198	0.297916383
5	0.183063313	0.20758413	0.261622237	0.296088618
6	0.183073896	0.20740496	0.260313514	0.294665403
7	0.183061971	0.20732091	0.259309147	0.293557306
8	0.183031162	0.20728849	0.258511937	0.292668868
9	0.182986288	0.20728807	0.25786263	0.291942826
10	0.182935159	0.20731767	0.257333044	0.291341235

We can observe from the above table that as the degree of the polynomial increases, the training rmse decreases as higher degree of polynomial provides more flexibility to fit the data.

The training rmse is decreasing as degree of polynomial increases for both gradient descent and stochastic gradient descent model.

As we increase the degree of the polynomial, there is a possibility that the testing rmse increases because of overfitting. But the above results show that the testing data rmse is also decreasing as the degree of polynomial increases. So, possibly there is no overfitting for this model.

V. After regularization

Minimum and training, validation and testing rmse error after introducing regularisation term:

Degree 1:

	Reg Const	GD_L1_train	GD_L1_validate	GD_L1_test	SGD_L1_train	SGD_L1_validate	SGD_L1_test
0	0.028347	0.174630	0.212234	0.210272	0.179315	0.221045	0.219962
1	0.093860	0.174635	0.212252	0.210290	0.204210	0.250096	0.251833
2	0.134364	0.174638	0.212262	0.210302	0.225983	0.273033	0.275938
3	0.255069	0.174646	0.212294	0.210337	0.269651	0.316792	0.321090
4	0.449491	0.174667	0.212365	0.210414	0.269261	0.316405	0.320690
5	0.495435	0.174670	0.212377	0.210427	0.269654	0.316795	0.321098
6	0.651593	0.174688	0.212438	0.210494	0.269759	0.316899	0.321200
7	0.763775	0.174697	0.212469	0.210527	0.269621	0.316759	0.321049
8	0.788723	0.174699	0.212475	0.210534	0.269629	0.316769	0.321077
9	0.847434	0.174704	0.212491	0.210552	0.269586	0.316731	0.321038

	Reg Const	GD_L2_train	GD_L2_validate	GD_L2_test	SGD_L2_train	SGD_L2_validate	SGD_L2_test
0	0.028347	0.174630	0.212234	0.210272	0.172302	0.210368	0.207983
1	0.093860	0.174635	0.212252	0.210290	0.174037	0.212768	0.210774
2	0.134364	0.174638	0.212262	0.210302	0.174871	0.213862	0.212025
3	0.255069	0.174646	0.212294	0.210337	0.176808	0.217458	0.215951
4	0.449491	0.174667	0.212365	0.210414	0.179998	0.222263	0.221150
5	0.495435	0.174670	0.212377	0.210427	0.180904	0.223485	0.222472
6	0.651593	0.174688	0.212438	0.210494	0.183792	0.227246	0.226517
7	0.763775	0.174697	0.212469	0.210527	0.186506	0.230592	0.230106
8	0.788723	0.174699	0.212475	0.210534	0.186922	0.231085	0.230637
9	0.847434	0.174704	0.212491	0.210552	0.187690	0.232012	0.231626

Degree 2:

	Reg Const	GD_L1_train	GD_L1_validate	GD_L1_test	SGD_L1_train	SGD_L1_validate	SGD_L1_test
0	0.028347	0.171270	0.208942	0.206685	0.178318	0.220806	0.219622
1	0.093860	0.171275	0.208964	0.206707	0.204483	0.250395	0.252150
2	0.134364	0.171285	0.208999	0.206744	0.226271	0.273324	0.276235
3	0.255069	0.171296	0.209039	0.206785	0.269395	0.316538	0.320833
4	0.449491	0.171320	0.209127	0.206876	0.269296	0.316437	0.320715
5	0.495435	0.171324	0.209142	0.206892	0.269416	0.316575	0.320874
6	0.651593	0.171339	0.209195	0.206947	0.269323	0.316474	0.320771
7	0.763775	0.171357	0.209256	0.207010	0.269591	0.316734	0.321040
8	0.788723	0.171360	0.209265	0.207019	0.269540	0.316671	0.320948
9	0.847434	0.171366	0.209285	0.207040	0.269691	0.316835	0.321154

	Reg Const	GD_L2_train	GD_L2_validate	GD_L2_test	SGD_L2_train	SGD_L2_validate	SGD_L2_test
0	0.028347	0.171270	0.208942	0.206685	0.170308	0.208211	0.205583
1	0.093860	0.171275	0.208964	0.206707	0.170885	0.209193	0.206774
2	0.134364	0.171285	0.208999	0.206744	0.171436	0.210173	0.207905
3	0.255069	0.171296	0.209039	0.206785	0.172407	0.212625	0.210435
4	0.449491	0.171320	0.209127	0.206876	0.175110	0.216901	0.215038
5	0.495435	0.171324	0.209142	0.206892	0.174984	0.216787	0.214891
6	0.651593	0.171339	0.209195	0.206947	0.177763	0.220682	0.219050
7	0.763775	0.171357	0.209256	0.207010	0.179349	0.222811	0.221298
8	0.788723	0.171360	0.209265	0.207019	0.180141	0.223832	0.222383
9	0.847434	0.171366	0.209285	0.207040	0.180509	0.224282	0.222868

Degree 3:

	Reg Const	GD_L1_train	GD_L1_validate	GD_L1_test	SGD_L1_train	SGD_L1_validate	SGD_L1_test
0	0.028347	0.170047	0.207905	0.205394	0.179216	0.222597	0.221494
1	0.093860	0.170053	0.207929	0.205419	0.204313	0.250221	0.251961
2	0.134364	0.170057	0.207945	0.205435	0.226280	0.273334	0.276254
3	0.255069	0.170074	0.208019	0.205511	0.269419	0.316566	0.320859
4	0.449491	0.170093	0.208094	0.205587	0.269605	0.316734	0.321024
5	0.495435	0.170098	0.208112	0.205605	0.269196	0.316338	0.320618
6	0.651593	0.170114	0.208173	0.205666	0.269423	0.316556	0.320856
7	0.763775	0.170125	0.208216	0.205711	0.269680	0.316821	0.321142
8	0.788723	0.170128	0.208226	0.205720	0.269617	0.316755	0.321036
9	0.847434	0.170134	0.208249	0.205744	0.269786	0.316922	0.321238

	Reg Const	GD_L2_train	GD_L2_validate	GD_L2_test	SGD_L2_train	SGD_L2_validate	SGD_L2_test
0	0.028347	0.170047	0.207905	0.205394	0.169913	0.207651	0.205062
1	0.093860	0.170053	0.207929	0.205419	0.170009	0.208098	0.205537
2	0.134364	0.170057	0.207945	0.205435	0.170167	0.208768	0.206210
3	0.255069	0.170074	0.208019	0.205511	0.170659	0.210094	0.207599
4	0.449491	0.170093	0.208094	0.205587	0.172781	0.214150	0.211857
5	0.495435	0.170098	0.208112	0.205605	0.173008	0.214545	0.212256
6	0.651593	0.170114	0.208173	0.205666	0.174381	0.216649	0.214487
7	0.763775	0.170125	0.208216	0.205711	0.175749	0.218611	0.216558
8	0.788723	0.170128	0.208226	0.205720	0.177282	0.220701	0.218769
9	0.847434	0.170134	0.208249	0.205744	0.177130	0.220502	0.218556

Degree 4:

	Reg Const	GD_L1_train	GD_L1_validate	GD_L1_test	SGD_L1_train	SGD_L1_validate	SGD_L1_test
0	0.028347	0.169868	0.207804	0.205130	0.178963	0.221721	0.220622
1	0.093860	0.169872	0.207830	0.205157	0.205100	0.251069	0.252865
2	0.134364	0.169875	0.207847	0.205173	0.226591	0.273655	0.276587
3	0.255069	0.169885	0.207895	0.205222	0.269625	0.316766	0.321066
4	0.449491	0.169900	0.207975	0.205303	0.269568	0.316722	0.321029
5	0.495435	0.169904	0.207993	0.205322	0.269674	0.316815	0.321107
6	0.651593	0.169918	0.208058	0.205387	0.269498	0.316652	0.320945
7	0.763775	0.169928	0.208104	0.205435	0.269423	0.316547	0.320824
8	0.788723	0.169931	0.208115	0.205445	0.269491	0.316624	0.320957
9	0.847434	0.169936	0.208139	0.205470	0.269901	0.317021	0.321330

	Reg Const	GD_L2_train	GD_L2_validate	GD_L2_test	SGD_L2_train	SGD_L2_validate	SGD_L2_test
0	0.028347	0.169868	0.207804	0.205130	0.169761	0.206804	0.204192
1	0.093860	0.169872	0.207830	0.205157	0.169940	0.208188	0.205560
2	0.134364	0.169875	0.207847	0.205173	0.169986	0.208367	0.205742
3	0.255069	0.169885	0.207895	0.205222	0.170431	0.209745	0.207165
4	0.449491	0.169900	0.207975	0.205303	0.171897	0.212812	0.210356
5	0.495435	0.169904	0.207993	0.205322	0.172002	0.213007	0.210554
6	0.651593	0.169918	0.208058	0.205387	0.173466	0.215425	0.213095
7	0.763775	0.169928	0.208104	0.205435	0.175720	0.218719	0.216582
8	0.788723	0.169931	0.208115	0.205445	0.175902	0.218963	0.216848
9	0.847434	0.169936	0.208139	0.205470	0.177309	0.220868	0.218869

Degree 5:

	Reg Const	GD_L1_train	GD_L1_validate	GD_L1_test	SGD_L1_train	SGD_L1_validate	SGD_L1_test
0	0.028347	0.170132	0.208025	0.205264	0.179206	0.222046	0.220981
1	0.093860	0.170134	0.208051	0.205290	0.204110	0.249988	0.251718
2	0.134364	0.170135	0.208067	0.205306	0.225751	0.272802	0.275698
3	0.255069	0.170133	0.208085	0.205325	0.268878	0.316027	0.320276
4	0.449491	0.170140	0.208163	0.205403	0.269767	0.316890	0.321212
5	0.495435	0.170142	0.208182	0.205422	0.269703	0.316840	0.321166
6	0.651593	0.170149	0.208246	0.205487	0.269760	0.316904	0.321197
7	0.763775	0.170148	0.208262	0.205504	0.269384	0.316522	0.320821
8	0.788723	0.170149	0.208273	0.205515	0.269334	0.316483	0.320799
9	0.847434	0.170152	0.208297	0.205539	0.269984	0.317098	0.321390

	Reg Const	GD_L2_train	GD_L2_validate	GD_L2_test	SGD_L2_train	SGD_L2_validate	SGD_L2_test
0	0.028347	0.170132	0.208025	0.205264	0.169943	0.207580	0.205135
1	0.093860	0.170134	0.208051	0.205290	0.170033	0.208102	0.205654
2	0.134364	0.170135	0.208067	0.205306	0.170140	0.208378	0.205784
3	0.255069	0.170133	0.208085	0.205325	0.170572	0.209849	0.207368
4	0.449491	0.170140	0.208163	0.205403	0.172153	0.213124	0.210647
5	0.495435	0.170142	0.208182	0.205422	0.172237	0.213286	0.210855
6	0.651593	0.170149	0.208246	0.205487	0.174435	0.216825	0.214550
7	0.763775	0.170148	0.208262	0.205504	0.174822	0.217394	0.215162
8	0.788723	0.170149	0.208273	0.205515	0.174941	0.217562	0.215368
9	0.847434	0.170152	0.208297	0.205539	0.175512	0.218387	0.216228

Degree 6:

	Reg Const	GD_L1_train	GD_L1_validate	GD_L1_test	SGD_L1_train	SGD_L1_validate	SGD_L1_test
0	0.028347	0.170480	0.208141	0.205395	0.179080	0.221859	0.220777
1	0.093860	0.170477	0.208164	0.205417	0.204102	0.249985	0.251707
2	0.134364	0.170469	0.208158	0.205413	0.226290	0.273342	0.276268
3	0.255069	0.170458	0.208182	0.205437	0.269681	0.316818	0.321131
4	0.449491	0.170438	0.208214	0.205471	0.269850	0.316957	0.321287
5	0.495435	0.170431	0.208213	0.205470	0.269356	0.316512	0.320802
6	0.651593	0.170414	0.208235	0.205495	0.269596	0.316756	0.321029
7	0.763775	0.170400	0.208243	0.205505	0.269520	0.316624	0.320921
8	0.788723	0.170399	0.208253	0.205514	0.269682	0.316817	0.321170
9	0.847434	0.170392	0.208258	0.205521	0.268874	0.316017	0.320268

	Reg Const	GD_L2_train	GD_L2_validate	GD_L2_test	SGD_L2_train	SGD_L2_validate	SGD_L2_test
0	0.028347	0.170480	0.208141	0.205395	0.170029	0.207796	0.205658
1	0.093860	0.170477	0.208164	0.205417	0.170256	0.208389	0.206011
2	0.134364	0.170469	0.208158	0.205413	0.170346	0.208894	0.206576
3	0.255069	0.170458	0.208182	0.205437	0.171202	0.211037	0.208654
4	0.449491	0.170438	0.208214	0.205471	0.171869	0.212339	0.209903
5	0.495435	0.170431	0.208213	0.205470	0.171833	0.212274	0.209845
6	0.651593	0.170414	0.208235	0.205495	0.173653	0.215464	0.213209
7	0.763775	0.170400	0.208243	0.205505	0.174316	0.216475	0.214231
8	0.788723	0.170399	0.208253	0.205514	0.175448	0.218144	0.216007
9	0.847434	0.170392	0.208258	0.205521	0.175616	0.218389	0.216262

Degree 7:

	Reg Const	GD_L1_train	GD_L1_validate	GD_L1_test	SGD_L1_train	SGD_L1_validate	SGD_L1_test
0	0.028347	0.170547	0.207912	0.205421	0.179118	0.221771	0.220708
1	0.093860	0.170531	0.207923	0.205428	0.204679	0.250609	0.252374
2	0.134364	0.170520	0.207926	0.205431	0.225683	0.272719	0.275619
3	0.255069	0.170481	0.207926	0.205433	0.269179	0.316327	0.320627
4	0.449491	0.170430	0.207945	0.205447	0.269318	0.316465	0.320782
5	0.495435	0.170414	0.207943	0.205447	0.269067	0.316244	0.320535
6	0.651593	0.170377	0.207964	0.205462	0.269591	0.316739	0.321012
7	0.763775	0.170346	0.207972	0.205468	0.269550	0.316643	0.320956
8	0.788723	0.170339	0.207974	0.205470	0.269185	0.316363	0.320683
9	0.847434	0.170324	0.207980	0.205474	0.269741	0.316904	0.321138

	Reg Const	GD_L2_train	GD_L2_validate	GD_L2_test	SGD_L2_train	SGD_L2_validate	SGD_L2_test
0	0.028347	0.170547	0.207912	0.205421	0.170147	0.208149	0.206085
1	0.093860	0.170531	0.207923	0.205428	0.170133	0.207872	0.205766
2	0.134364	0.170520	0.207926	0.205431	0.170334	0.208715	0.206589
3	0.255069	0.170481	0.207926	0.205433	0.171346	0.210880	0.208465
4	0.449491	0.170430	0.207945	0.205447	0.172094	0.212569	0.210282
5	0.495435	0.170414	0.207943	0.205447	0.173234	0.214527	0.212251
6	0.651593	0.170377	0.207964	0.205462	0.174126	0.215984	0.213808
7	0.763775	0.170346	0.207972	0.205468	0.174555	0.216564	0.214353
8	0.788723	0.170339	0.207974	0.205470	0.175855	0.218591	0.216585
9	0.847434	0.170324	0.207980	0.205474	0.175918	0.218664	0.216661

Degree 8:

	Reg Const	GD_L1_train	GD_L1_validate	GD_L1_test	SGD_L1_train	SGD_L1_validate	SGD_L1_test
0	0.028347	0.170468	0.207809	0.205593	0.179468	0.222867	0.221802
1	0.093860	0.170443	0.207810	0.205585	0.204549	0.250473	0.252239
2	0.134364	0.170430	0.207813	0.205582	0.225304	0.272344	0.275189
3	0.255069	0.170391	0.207823	0.205573	0.269367	0.316508	0.320817
4	0.449491	0.170332	0.207843	0.205561	0.269187	0.316344	0.320615
5	0.495435	0.170320	0.207850	0.205559	0.269445	0.316576	0.320885
6	0.651593	0.170279	0.207872	0.205553	0.269823	0.316962	0.321296
7	0.763775	0.170246	0.207882	0.205545	0.269500	0.316620	0.320905
8	0.788723	0.170239	0.207884	0.205542	0.268571	0.315747	0.320007
9	0.847434	0.170232	0.207903	0.205549	0.269996	0.317169	0.321411

	Reg Const	GD_L2_train	GD_L2_validate	GD_L2_test	SGD_L2_train	SGD_L2_validate	SGD_L2_test
0	0.028347	0.170468	0.207809	0.205593	0.169892	0.207471	0.205593
1	0.093860	0.170443	0.207810	0.205585	0.170413	0.208826	0.206782
2	0.134364	0.170430	0.207813	0.205582	0.170440	0.208723	0.206620
3	0.255069	0.170391	0.207823	0.205573	0.171177	0.210540	0.208333
4	0.449491	0.170332	0.207843	0.205561	0.172190	0.212624	0.210444
5	0.495435	0.170320	0.207850	0.205559	0.171977	0.212005	0.209739
6	0.651593	0.170279	0.207872	0.205553	0.174185	0.215775	0.213627
7	0.763775	0.170246	0.207882	0.205545	0.175134	0.217288	0.215247
8	0.788723	0.170239	0.207884	0.205542	0.175514	0.217922	0.215949
9	0.847434	0.170232	0.207903	0.205549	0.175625	0.217990	0.215967

Degree 9:

	Reg Const	GD_L1_train	GD_L1_validate	GD_L1_test	SGD_L1_train	SGD_L1_validate	SGD_L1_test
0	0.028347	0.170380	0.207751	0.205715	0.179254	0.222101	0.221039
1	0.093860	0.170359	0.207759	0.205701	0.204317	0.250221	0.251961
2	0.134364	0.170344	0.207761	0.205690	0.226134	0.273182	0.276090
3	0.255069	0.170301	0.207768	0.205659	0.269375	0.316517	0.320827
4	0.449491	0.170238	0.207784	0.205613	0.269370	0.316508	0.320774
5	0.495435	0.170230	0.207797	0.205610	0.269536	0.316696	0.321010
6	0.651593	0.170196	0.207827	0.205594	0.269562	0.316686	0.320973
7	0.763775	0.170186	0.207866	0.205597	0.269690	0.316776	0.321096
8	0.788723	0.170183	0.207873	0.205597	0.269470	0.316651	0.320962
9	0.847434	0.170167	0.207875	0.205586	0.269709	0.316893	0.321155

	Reg Const	GD_L2_train	GD_L2_validate	GD_L2_test	SGD_L2_train	SGD_L2_validate	SGD_L2_test
0	0.028347	0.170380	0.207751	0.205715	0.169942	0.207649	0.205809
1	0.093860	0.170359	0.207759	0.205701	0.170540	0.209185	0.207187
2	0.134364	0.170344	0.207761	0.205690	0.170667	0.209632	0.207622
3	0.255069	0.170301	0.207768	0.205659	0.171095	0.210500	0.208442
4	0.449491	0.170238	0.207784	0.205613	0.172180	0.212488	0.210369
5	0.495435	0.170230	0.207797	0.205610	0.172723	0.213176	0.210994
6	0.651593	0.170196	0.207827	0.205594	0.174705	0.216705	0.214783
7	0.763775	0.170186	0.207866	0.205597	0.174544	0.216231	0.214208
8	0.788723	0.170183	0.207873	0.205597	0.175433	0.217632	0.215704
9	0.847434	0.170167	0.207875	0.205586	0.175949	0.218391	0.216511

Degree 10:

	Reg Const	GD_L1_train	GD_L1_validate	GD_L1_test
0	0.028347	0.170298	0.207716	0.205782
1	0.093860	0.170281	0.207729	0.205762
2	0.134364	0.170266	0.207730	0.205743
3	0.255069	0.170239	0.207756	0.205709
4	0.449491	0.170197	0.207792	0.205660
5	0.495435	0.170195	0.207810	0.205659
6	0.651593	0.170175	0.207846	0.205638
7	0.763775	0.170158	0.207866	0.205620
8	0.788723	0.170151	0.207864	0.205612
9	0.847434	0.170151	0.207887	0.205616

	Reg Const	GD_L2_train	GD_L2_validate	GD_L2_test
0	0.028347	0.170298	0.207716	0.205782
1	0.093860	0.170281	0.207729	0.205762
2	0.134364	0.170266	0.207730	0.205743
3	0.255069	0.170239	0.207756	0.205709
4	0.449491	0.170197	0.207792	0.205660
5	0.495435	0.170195	0.207810	0.205659
6	0.651593	0.170175	0.207846	0.205638
7	0.763775	0.170158	0.207866	0.205620
8	0.788723	0.170151	0.207864	0.205612
9	0.847434	0.170151	0.207887	0.205616

SGD_L1_train	SGD_L1_validate	SGD_L1_test
0.178697	0.221079	0.219965
0.204415	0.250324	0.252073
0.226800	0.273869	0.276827
0.269258	0.316409	0.320693
0.269698	0.316836	0.321112
0.269124	0.316286	0.320589
0.269628	0.316753	0.321029
0.269593	0.316728	0.321027
0.269390	0.316557	0.320874
0.269374	0.316512	0.320808

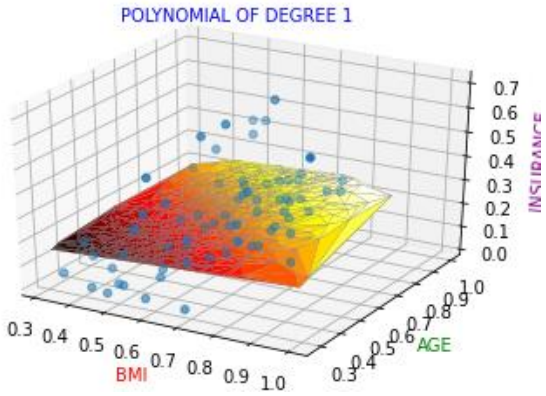
SGD_L2_train	SGD_L2_validate	SGD_L2_test
0.170029	0.208125	0.206281
0.170019	0.208126	0.206106
0.170312	0.208733	0.206750
0.171537	0.211322	0.209357
0.173105	0.214131	0.212176
0.172768	0.213544	0.211569
0.174455	0.216031	0.214084
0.175149	0.217098	0.215241
0.175473	0.217479	0.215650
0.175850	0.218150	0.216353

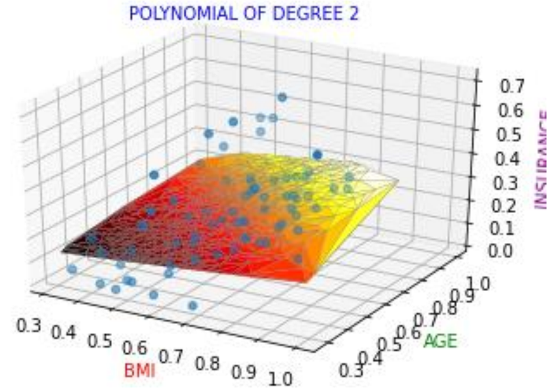
From the above observations, we can see that ridge regression gives lesser rmse for both gradient descent and stochastic gradient descent for all degree polynomials (1-10).

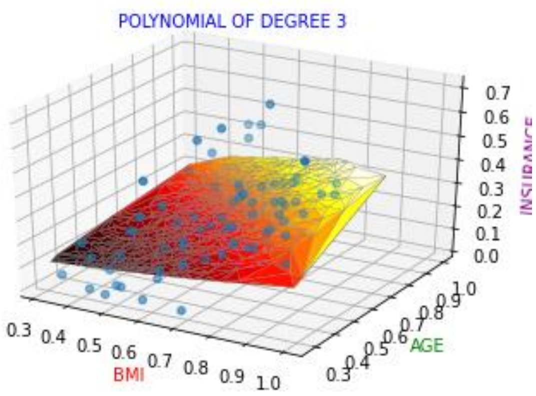
This is because we have only 2 features for our model and Lasso regression yields better results for data having large number of features.

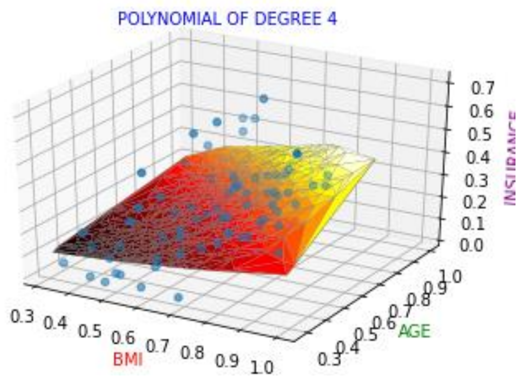
Also, as we try to fit the data into polynomial of degree 10, the validation error increases. This can be an indicator that polynomial of degree 10 is an overfit.

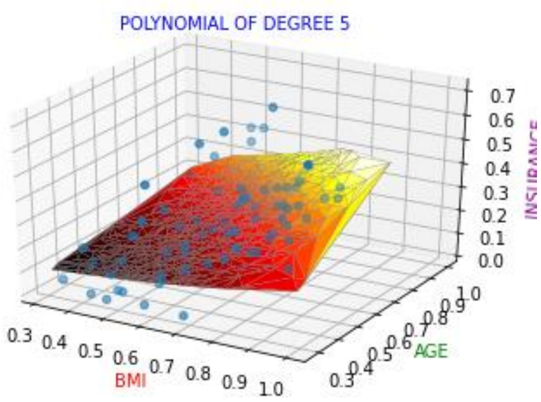
VI. Surface plots

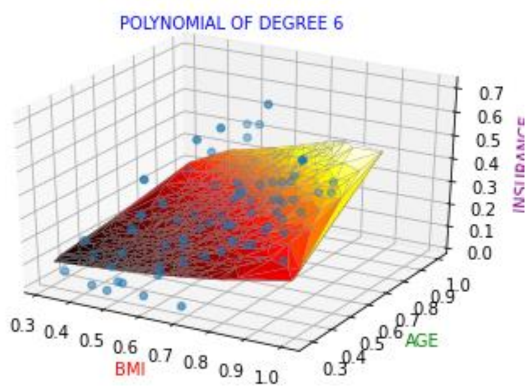


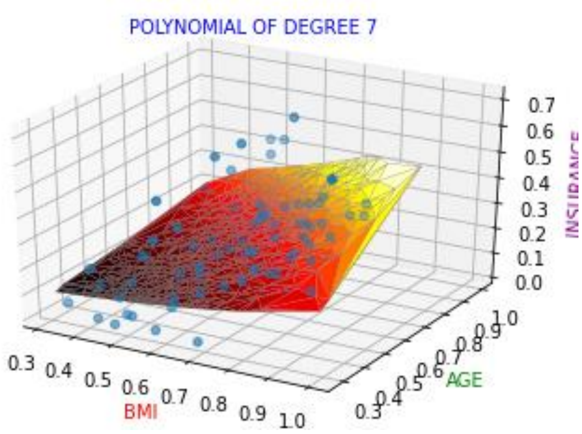


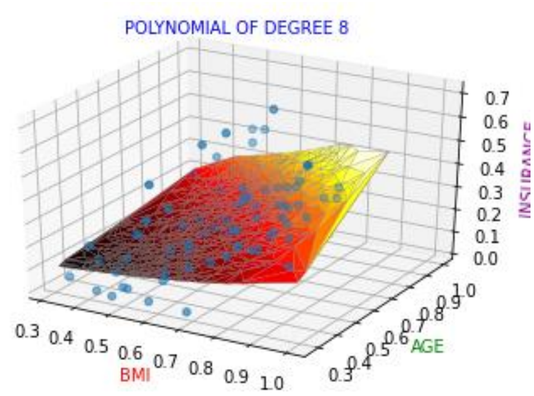


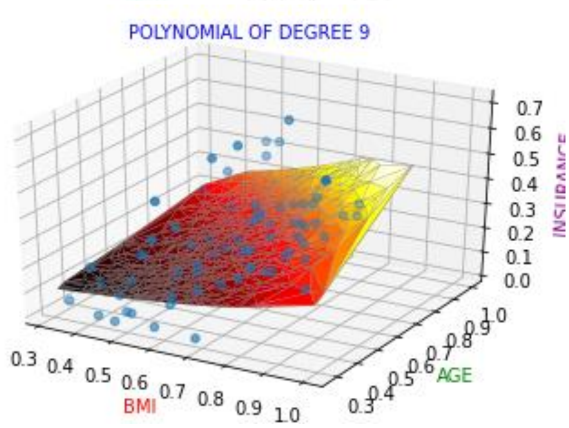


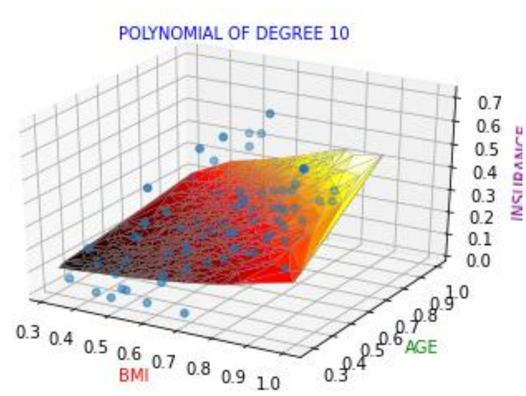












VII. Questions to ponder

Q. What happens to the training and testing error as polynomials of higher degree are used for prediction?

A. Training error will keep on reducing as we keep using a higher polynomial as we get higher degrees of freedom. However, for testing data we can't say conclusively as it depends on the polynomial models the data well. However, for higher degree polynomials the testing error tends to be high as overfitting comes into play due to high degree of freedom for the weights.

Q. Does a single global minimum exist for Polynomial Regression as well? If yes, justify.

A. Yes. This is because irrespective of the degree of the polynomial the error function always remains quadratic and hence it's minima can be found easily. Any polynomial regression equation can be solved to obtain the final weights to minimize the error by differentiating w.r.t the weights w0, w1…, w11, w12,,wNN.

Q. Which form of regularization curbs overfitting better in your case? Can you think of a case when Lasso regularization works better than Ridge?

A. Lasso tends to do well if there are a small number of significant parameters and the others are close to zero (ergo: when only a few predictors influence the response).
Ridge works well if there are many large parameters of about the same value (ergo: when most predictors impact the response).
However, in practice, we don't know the true parameter values, so the previous two points are somewhat theoretical.

Q. How does the regularization parameter affect the regularization process and weights? What would happen if a higher value for λ (> 2) was chosen?

A. Larger the value of λ the smaller the weights have to be (which will introduce a bias and the weights won't model the data very well). Large value of λ will make minimizing the sum of squares very difficult. λ has to be chosen very carefully as it can't be too large or too small.
If λ is too large, then all importance is given to stopping growth of weights and none to minimizing the cost function itself.

Q. Regularization is necessary when you have a large number of features but limited training instances. Do you agree with this statement?

A. Yes, we agree with this statement as, for smaller training instances larger number of features gives a larger degree of freedom and will lead to significant overfitting and lead to poor outputs with testing data. Thus, to counter overfitting regularization is absolutely needed.

Q. If you are provided with D original features and are asked to generate new matured features of degree N, how many such new matured features will you be able to generate? Answer in terms of N and D.

A. Number of features for (D,N)= (D+N)C(N) where D is the number of the original features, N is the degree of the polynomial.

Q. What is bias-variance trade off and how does it relate to overfitting and regularization.

A. If our model is too simple and has very few parameters then it may have high bias and low variance. On the other hand, if our model has large number of parameters then it's going to have high variance and low bias. So, we need to find the right/good balance without overfitting and underfitting the data.

This trade-off in complexity is why there is a trade-off between bias and variance. An algorithm can't be more complex and less complex at the same time.

Linear machine learning algorithms often have a high bias but a low variance.

Nonlinear machine learning algorithms often have a low bias but a high variance.

The parameterization of machine learning algorithms is often a battle to balance out bias and variance.