Assignment 1 Report

1. Obtaining Dataset

In this assignment, I used the **Abalone** datasets from **UCI Dataset**.

Variable Name	Role	Type
sex	Feature	Categorical
length	Feature	Continuous
diameter	Feature	Continuous
height	Feature	Continuous
whole_weight	Feature	Continuous
shucked_weight	Feature	Continuous
viscera_weight	Feature	Continuous
shell_weight	Feature	Continuous
rings	Target	Integer

2. Preprocessing Dataset

The data in the dataset do have **missing value**. There are one column (sex) with categorical type that we need to convert into numerical type for regression.

Below are the correlation for each features relative to our target features (ring):

Feature	Correlation
rings	1.000000
shell_weight	0.627574
diameter	0.574660
height	0.557467
length	0.556720
whole_weight	0.540390
viscera_weight	0.503819
shucked_weight	0.420884
sex	-0.351822

3. Regression Model

3.1. Gradient Descent Model from Scratch

We tried different paramters for the Gradient Descent function and have the following results:

Learning Rate	Threshold	Max Iterations	R^2
0.001	1e-06	5000	0.23017070049886923
0.001	1e-06	10000	0.2791417054214471
0.001	1e-06	50000	0.336286923008177
0.001	1e-05	5000	0.2208798731742413
0.001	1e-05	10000	0.2784637773035321
0.001	1e-05	50000	0.3501774905991951
0.001	0.0001	5000	0.2464217582736833
0.001	0.0001	10000	0.2568013181328638
0.001	0.0001	50000	0.26583345756015364
0.01	1e-06	5000	0.3364211753318098
0.01	1e-06	10000	0.3934602546320263
0.01	1e-06	50000	0.49637901974386556
0.01	1e-05	5000	0.3320179207540752
0.01	1e-05	10000	0.38750138549983637
0.01	1e-05	50000	0.49644005625802123
0.01	0.0001	5000	0.343475285682342
0.01	0.0001	10000	0.39507484198769083
0.01	0.0001	50000	0.49610393712590006
0.1	1e-06	5000	0.4964161913125126
0.1	1e-06	10000	0.5040860755130276
0.1	1e-06	50000	0.505831461159481
0.1	1e-05	5000	0.4964891321272551
0.1	1e-05	10000	0.503906428784517
0.1	1e-05	50000	0.5057417979549297
0.1	0.0001	5000	0.49627959672841193
0.1	0.0001	10000	0.5038119138938579
0.1	0.0001	50000	0.506123864467745

From the table we noticed that our best parameters for the model is: `Learning rate=0.1, Threshold=0.0001, Max iterations=50000` with the best R2 score of `0.506123864467745`

In my opinion, I think that this is the best that linear regression can do for this datasets. The problem that we have a lower R2 score might be that the datasets deviate alot from linearity, therefore using linear model in such cases yield us lower results.

3.2. Using SGDRegressor Model

We used Grid Search to try different parameters from SGDRegressor Model.

Below is the results we have:

Alpha	Learning Rate	Tolerance	R^2 Mean	R^2 Std
0.0001	constant	0.001	0.5188	0.0441
0.0001	constant	0.0001	0.5209	0.0391
0.0001	constant	1e-05	0.5046	0.0285
0.0001	constant	0.001	0.5104	0.0420
0.0001	constant	0.0001	0.5243	0.0426
0.0001	constant	1e-05	0.5124	0.0402
0.0001	constant	0.001	0.5150	0.0423
0.0001	constant	0.0001	0.5207	0.0421
0.0001	constant	1e-05	0.5187	0.0491
0.0001	optimal	0.001	-2369923824.7547	2257229377.7960
0.0001	optimal	0.0001	-908294991.2302	511610229.4383
0.0001	optimal	1e-05	-2093479908.4644	2033191056.1277
0.0001	optimal	0.001	-100356490.3375	188376658.7237
0.0001	optimal	0.0001	-48346896.4304	36469190.2040
0.0001	optimal	1e-05	-73115180.8881	97748647.4528
0.0001	optimal	0.001	-4750522.8980	4832926.3089
0.0001	optimal	0.0001	-2391503.4616	4297237.9358
0.0001	optimal	1e-05	-2893713.3706	3993490.3311
0.0001	invscaling	0.001	0.5038	0.0345
0.0001	invscaling	0.0001	0.5097	0.0334
0.0001	invscaling	1e-05	0.4959	0.0253
0.0001	invscaling	0.001	0.5042	0.0325
0.0001	invscaling	0.0001	0.5066	0.0301
0.0001	invscaling	1e-05	0.5105	0.0320
0.0001	invscaling	0.001	0.4991	0.0324
0.0001	invscaling	0.0001	0.5078	0.0355
0.0001	invscaling	1e-05	0.5096	0.0332
0.0001	adaptive	0.001	0.5282	0.0431
0.0001	adaptive	0.0001	0.5269	0.0422
0.0001	adaptive	1e-05	0.5272	0.0421
0.0001	adaptive	0.001	0.5275	0.0429
0.0001	adaptive	0.0001	0.5272	0.0414
0.0001	adaptive	1e-05	0.5280	0.0423
0.0001	adaptive	0.001	0.5276	0.0430
0.0001	adaptive	0.0001	0.5275	0.0421
0.0001	adaptive	1e-05	0.5275	0.0424
0.001	constant	0.001	0.4928	0.0485
0.001	constant	0.0001	0.5117	0.0388
0.001	constant	1e-05	0.5053	0.0330
0.001	constant	0.001	0.4912	0.0469
0.001	constant	0.0001	0.4988	0.0383
0.001	constant	1e-05	0.5118	0.0406
0.001	constant	0.001	0.5012	0.0381

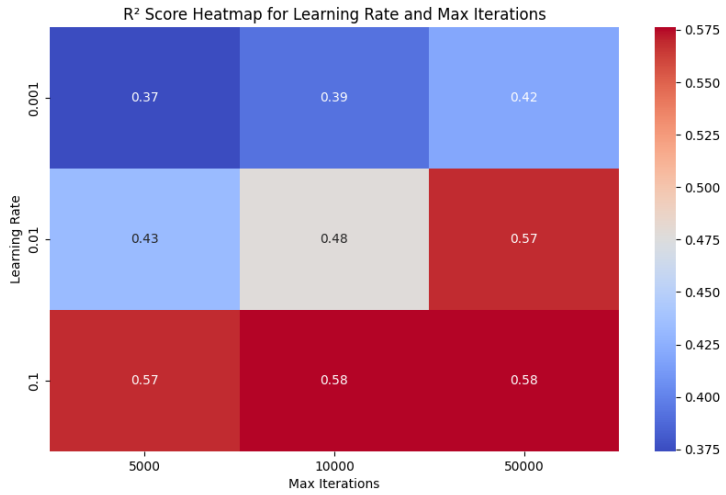
Alpha	Learning Rate	Tolerance	R^2 Mean	R^2 Std
0.001	constant	0.0001	0.5010	0.0389
0.001	constant	1e-05	0.4998	0.0330
0.001	optimal	0.001	-191273626673.9283	182663769550.2284
0.001	optimal	0.0001	-64917194708.3252	60220270334.2418
0.001	optimal	1e-05	-68511761582.4003	54277802781.6777
0.001	optimal	0.001	-37994925056.3963	44026569495.3089
0.001	optimal	0.0001	-13386798514.4134	8463383768.0873
0.001	optimal	1e-05	-21478030605.9519	29733443749.2251
0.001	optimal	0.001	-8642685741.1243	11311595684.9233
0.001	optimal	0.0001	-4810665890.4899	2964356255.4699
0.001	optimal	1e-05	-7804047658.5065	10724951870.0598
0.001	invscaling	0.001	0.4838	0.0226
0.001	invscaling	0.0001	0.4944	0.0301
0.001	invscaling	1e-05	0.4949	0.0274
0.001	invscaling	0.001	0.4884	0.0256
0.001	invscaling	0.0001	0.4897	0.0254
0.001	invscaling	1e-05	0.4877	0.0264
0.001	invscaling	0.001	0.4922	0.0269
0.001	invscaling	0.0001	0.4886	0.0232
0.001	invscaling	1e-05	0.4961	0.0318
0.001	adaptive	0.001	0.5174	0.0357
0.001	adaptive	0.0001	0.5172	0.0358
0.001	adaptive	1e-05	0.5176	0.0357
0.001	adaptive	0.001	0.5172	0.0361
0.001	adaptive	0.0001	0.5175	0.0360
0.001	adaptive	1e-05	0.5177	0.0358
0.001	adaptive	0.001	0.5172	0.0357
0.001	adaptive	0.0001	0.5174	0.0357
0.001	adaptive	1e-05	0.5177	0.0366
0.01	constant	0.001	0.3775	0.0337
0.01	constant	0.0001	0.3768	0.0432
0.01	constant	1e-05	0.4002	0.0212
0.01	constant	0.001	0.4032	0.0152
0.01	constant	0.0001	0.3933	0.0171
0.01	constant	1e-05	0.4034	0.0198
0.01	constant	0.001	0.3965	0.0249
0.01	constant	0.0001	0.3924	0.0166
0.01	constant	1e-05	0.4003	0.0256
0.01	optimal	0.001	-64761568869.2573	59360998636.9355
0.01	optimal	0.0001	-73791980832.5152	33265507162.6219
0.01	optimal	1e-05	-217466359223.7168	138793891072.1992
0.01	optimal	0.001	-27834432733.8594	19968580739.3203
0.01	optimal	0.0001	-13472335545.8101	17376616114.3561

Alpha	Learning Rate	Tolerance	R^2 Mean	R^2 Std
0.01	optimal	1e-05	-67183933429.1207	56772917061.6034
0.01	optimal	0.001	-24990623976.5404	30675641566.4915
0.01	optimal	0.0001	-14457642153.4381	9123249636.9311
0.01	optimal	1e-05	-24372387082.7489	21166982840.7638
0.01	invscaling	0.001	0.3991	0.0189
0.01	invscaling	0.0001	0.3971	0.0219
0.01	invscaling	1e-05	0.3993	0.0186
0.01	invscaling	0.001	0.4008	0.0175
0.01	invscaling	0.0001	0.3974	0.0173
0.01	invscaling	1e-05	0.3977	0.0174
0.01	invscaling	0.001	0.3984	0.0176
0.01	invscaling	0.0001	0.3991	0.0176
0.01	invscaling	1e-05	0.4022	0.0179
0.01	adaptive	0.001	0.4090	0.0195
0.01	adaptive	0.0001	0.4091	0.0194
0.01	adaptive	1e-05	0.4088	0.0196
0.01	adaptive	0.001	0.4088	0.0195
0.01	adaptive	0.0001	0.4090	0.0195
0.01	adaptive	1e-05	0.4089	0.0195
0.01	adaptive	0.001	0.4090	0.0195
0.01	adaptive	0.0001	0.4089	0.0195
0.01	adaptive	1e-05	0.4089	0.0195

We notice that `learning_rate=optimal` is not good at all because it yield us negative R2 score. The best parameter that we found is `'alpha': 0.0001, 'learning_rate': 'adaptive', 'max_iter': 1000, 'tol': 0.001` with R2 score of `0.5282258489110061`.

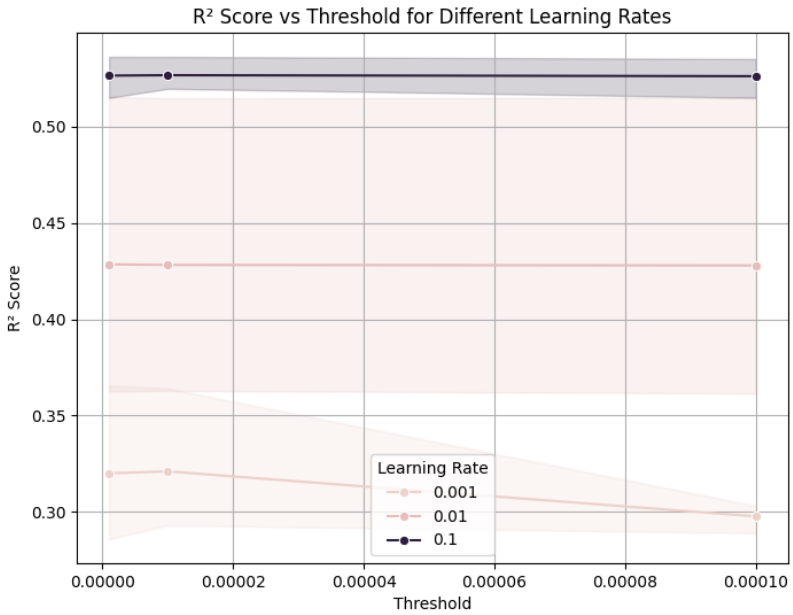
Comparing this model to the model that we built from scratch, we notice that they both have similar performance and yield similar R2 results. However, the SGDRegressor needed less iterations and tolerance in order to achieve good results compared to the model that we built. This shows that there might be more optimization that we can do to improve our linear regression model.

From the table above, we can generate the following graphs:



In this graph, we notice that the inverse relationship of learning rate. It proves that we do not choose the smallest learning rate but the optimal one.

In this graph, we notice that the R2 score starts going constant starting from 10,000 iterations for our optimal learning rate



In this graph, the r2 does not change even when the threshold goes smaller than the optimal point. This means that once we reach the optimal point of a parameter, decreasing or increasing it further does not enhance or change the results.

4. Conclusions

In my conclusion, even if the R2 score seems okay, there is a much better approach than pure linear regression to solve this problem. Having a slightly lower R2 score indicates that the dataset might not fit in a linear model and that we should explore other options that aligns more with the nature of the dataset itself.