M,D&L Assignment-1 Report

05-02-2021

Palash Sharma
2019101082

Aryan Jain
2019101056

Task 1

LinearRegression fits a linear model with coefficients w = (w1, ..., wp) to minimize the residual sum of squares between the observed targets in the dataset, and the targets predicted by the linear approximation. The function LinearRegression.fit() does the specific task of trying out different values of coefficients before settling for the best fitting set of weights. It takes training data as a set of arguments and can be used for supervised as well as unsupervised learning. It is interesting to note that polynomial regression still falls under the linear model since one variable associates to one predicted variable. The method involves using derivatives of MSE(Mean Squared Error) to decide whether to increase or decrease the weights and a learning rate decided by the user determines the size of steps that the model takes to reach a conclusion. It is advisable that learning rate be not too small or large to give the optimal results in a reasonable amount of time.

Task 2

	Degree	Variance	Bias^2	Irreducible Error
0	1	[25999.093009987824]	[1001682.5006488453]	[-1.1641532182693481e-10]
1	2	[39105.833813269084]	[953836.1284241962]	[8.149072527885437e-10]
2	3	[56095.89320974847]	[9533.344066289577]	[2.801243681460619e-10]
3	4	[114907.29153006151]	[10588.29843702177]	[8.203642209991813e-10]
4	5	[151434.02783423715]	[9952.445160291447]	[-5.802576197311282e-10]
5	6	[174226.7458271964]	[9999.198129961162]	[4.729372449219227e-11]
6	7	[198849.6324516319]	[10425.934603741476]	[1.7826096154749393e-10]
7	8	[221547.4033618678]	[10999.522314961298]	[1.2005330063402653e-10]
8	9	[232479.58321112586]	[11449.147224149043]	[4.18367562815547e-11]
9	10	[233184.61082255776]	[14482.835786796204]	[-1.7462298274040222e-10]
10	11	[238591.26694883226]	[12413.174767704015]	[2.546585164964199e-11]
11	12	[218974.84218720612]	[30461.025334682858]	[5.056790541857481e-10]
12	13	[233068.3055957974]	[16061.638932883012]	[-3.4924596548080444e-10]
13	14	[212545.42818291223]	[39231.05485150342]	[-5.093170329928398e-10]
14	15	[221714.93435395617]	[62928.35277764195]	[4.3655745685100555e-11]
15	16	[239358.04420893276]	[69821.4113264154]	[2.9103830456733704e-10]
16	17	[242993.9622762862]	[115239.38837579123]	[2.9103830456733704e-11]
17	18	[269048.95988195285]	[120580.64334333695]	[-7.275957614183426e-11]
18	19	[270102.6712632651]	[191979.61420719136]	[2.6193447411060333e-10]
19	20	[299019.40609669394]	[197285.9213929445]	[1.0186340659856796e-09]

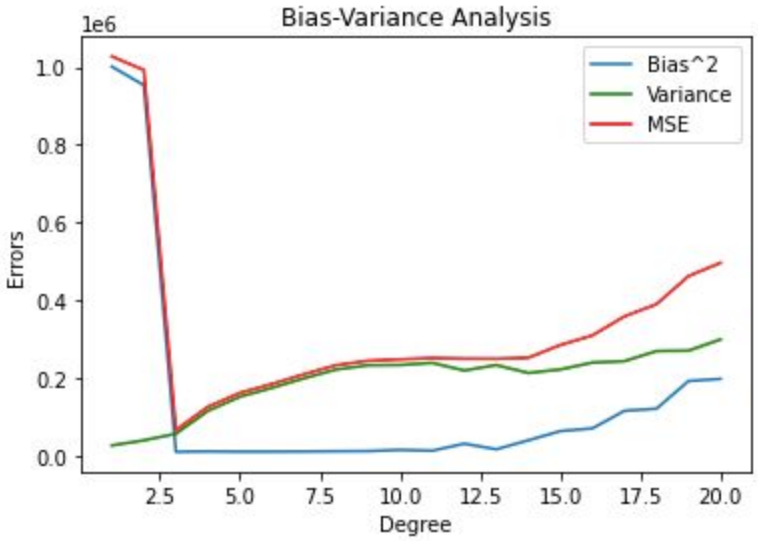
Bias is a measure of the distance between the reality and how the model perceives it. Bias is supposed to be least where the best fit has occurred. This is visible in our results where we see the lowest bias for degree 3 which means it is also the best fit model and the bias increases as we go away from the best fit meaning the models may have under/over fit with the training data.

Variance is a measure of the convulsions in a trained model which is measured with respect to the mean of all the values. Variance generally increases with the model complexity. This same thing can also be observed from our results where variance mostly increases with model complexity.

Task 3

Visibly, our results shows that the irreducible errors have stayed ridiculously less in value and hovering both sides of the zero which shows us that the errors have stayed the same over all model complexities and this correct conceptually as well since it shows the noise in dataset, be it training or test datasets. Irreducible error cannot be reduced by creating good models. **The negative values for irreducible error arise due to floating point errors.**

Task 4



From the graph above, we can infer that degree 3 model gives the best bias variance tradeoff as earlier models seem to severely underfit whereas the latter models do give considerable bias values but the variance too shoots up which makes for a bad tradeoff therefore an overfit model. Dataset seems minimal on noise since the irreducible error always stays close to zero. The nature of the data is best described by a degree three model.