Kamalini Pradhan

Date : 07/19/2024

Purpose

This assignment is to explore the use of scikit-learn package to build different regression models

GitHub Repo : [S573647/AMLLab4 (github.com)](https://github.com/S573647/AMLLab4)

*Assignment on Building a Regression Model*

Table of Contents

[Progress 1 - Screen shot pipeline ready to replace classifier with regression model. 2](#_Toc172323432)

[Progress 2 - Screen shot performance of model on training and test sets. 2](#_Toc172323433)

[Progress 3 - Analysis 2](#_Toc172323434)

[Compare performance and argue if model is underfitting the training set. 2](#_Toc172323435)

[Progress 4 - Screen shot linear fit 3](#_Toc172323436)

[Submission 5 - Analysis 3](#_Toc172323437)

[a) Will more training data improve the fit? 3](#_Toc172323438)

[b) Is the model under or over fitting? 3](#_Toc172323439)

[Submission 6 - Explain 3](#_Toc172323440)

[Did adding age improve the model? 3](#_Toc172323441)

[Propose an explanation for the results. 3](#_Toc172323442)

[Progress 7 - Scatter graph 3](#_Toc172323443)

[Scatter graph with cubic model displayed. 3](#_Toc172323444)

[Propose an explanation. 3](#_Toc172323445)

[Submission 8 - Explain 3](#_Toc172323446)

[a) Does the polynomial fit do better? 3](#_Toc172323447)

[b) Where does it fit the best? 3](#_Toc172323448)

[Progress 9 - Screen shot of scatter plot with degree 8 polynomial fit 3](#_Toc172323449)

[Submission 10 - Explain 4](#_Toc172323450)

[a) Compare degree 3 and degree 8 fits 4](#_Toc172323451)

[b) Is the increase in performance big enough to justify the degree 8 fit? 4](#_Toc172323452)

[Progress 11 - Screen shot of scatter plot with degree 8 Elastic net. 4](#_Toc172323453)

[Submission 12- Explain 4](#_Toc172323454)

[Compare coefficients for regular and elastic net for degree 8. 4](#_Toc172323455)

[Which ones have been reduced in elastic net? 4](#_Toc172323456)

# Progress 1 - Screen shot pipeline ready to replace classifier with regression model.

A screenshot of a computer

Description automatically generated

# Progress 2 - Screen shot performance of model on training and test sets.

A screenshot of a computer

Description automatically generated

A screenshot of a computer program

Description automatically generated

# Progress 3 - Analysis

## Compare performance and argue if model is underfitting the training set.

A screenshot of a computer

Description automatically generatedThe model does not appear to be underft. The linear regression model shows high R² scores of 0.89 for the training set and 0.86 for the testing set, indicating that it explains a substantial proportion of the variance in the dependent variable for both datasets. The RMSE values are 5.02 for the training set and 4.83 for the testing set, which are relatively low and comparable, suggesting that the model's predictions are close to the actual values. The slight difference in performance between the training and testing sets implies that the model generalizes well to new data. If the model were underfit, we would expect significantly lower R² scores and higher RMSE values, along with poor performance on both datasets. Therefore, the model effectively captures the underlying patterns in the data without underfitting.

# Progress 4 - Screen shot linear fit

A screen shot of a computer

Description automatically generated

A screen shot of a graph

Description automatically generated

# Submission 5 - Analysis

## Will more training data improve the fit?

The quality and relevance of the additional data matter. More data that is similar to the existing data can improve the model's performance. If the new data is noisy or not representative of the real-world scenario, it may not help or could even worsen the model’s performance.

## Is the model under or over fitting?

No because the performance on the training and testing sets is quite close, indicating that the model does not suffer from under or over fitting.

# Submission 6 - Explain

## Did adding age improve the model?

Yes, because RMSE is lesser compared to the height as an input.

## Propose an explanation for the results.

A black screen with white text

Description automatically generated A black and white screen with white text

Description automatically generated

R2 should be higher and RMSE should be lower for better performance. Comparing both the tables, R2 is the same. However, RMSE for the second table is lower than RMSE for the first table. Hence, model in the second table performs better.

# Progress 7 - Scatter graph

## Scatter graph with cubic model displayed.

A screen shot of a graph

Description automatically generated

## Propose an explanation.

The polynomial regression for power 3 performs better than the linear regression as you can see that the polynomial regression curve is better fit. Also, the performance metrics can be seen in the following table, where R2 is higher and RMSE is lower.  
A black and white screen

Description automatically generated

# Submission 8 - Explain

## Does the polynomial fit do better?

Yes

## b) Where does it fit the best?

Polynomial regression with power 8 is better because R2 is higher than the polynomial regression with power 3, and RMSE is lower as you can see in the following tables.

Polynomial regression with power 3:

A black and white screen

Description automatically generated

Polynomial regression with power 8: A screen shot of a black background

Description automatically generated

# Progress 9 - Screen shot of scatter plot with degree 8 polynomial fit

A screen shot of a computer

Description automatically generated

# Submission 10 - Explain

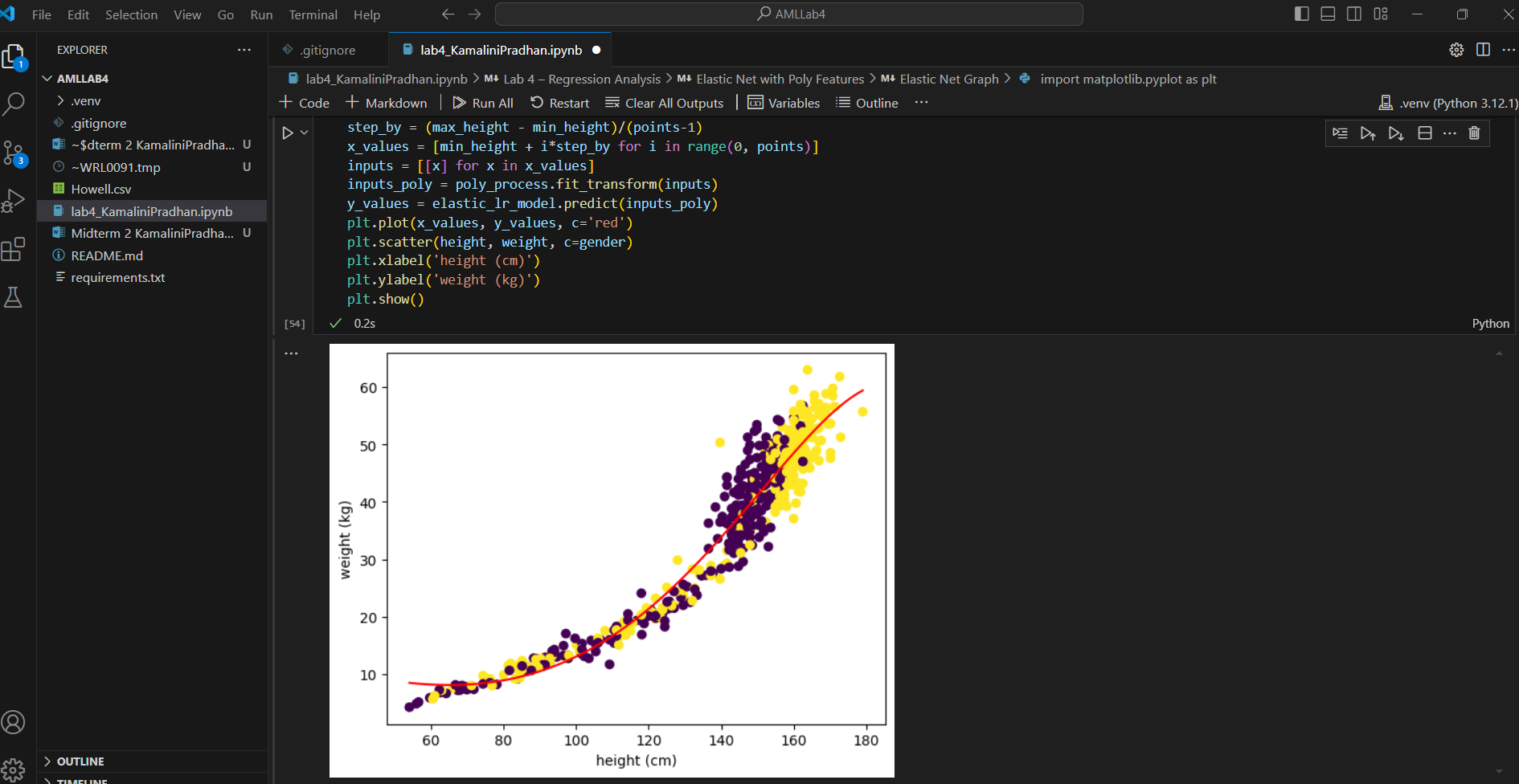
## Compare degree 3 and degree 8 fits

Polynomial regression with degree 8 has better performance than the polynomial regression with degree 3 because it has higher R2 and lower RMSE. Also, it is much better fit as you can see in the above picture.

## Is the increase in performance big enough to justify the degree 8 fit?

Yes, although there is no much difference in R2, there is substantial difference in RMSE, which justify that polynomial regression with power 8 performs better.

# Progress 11 - Screen shot of scatter plot with degree 8 Elastic net.



# Submission 12- Explain

## Compare coefficients for regular and elastic net for degree 8.

Polynomial with degree 8:

Coefficients:



Elastic net with degree 8:

Coefficients:



## 

## Which ones have been reduced in elastic net?

|  |  |  |  |
| --- | --- | --- | --- |
| Coefficient Number | Coefficient for polynomial with degree 8 | Relation | Coefficient for Elastic net with degree 8 |
| 1 | -4.87856869e-07 | < | -2.22044117e-01 |
| 2 | -2.87216867e-05 | < | 5.22052404e-04 |
| 3 | -7.94873980e-04 | < | 8.14877047e-06 |
| 4 | 3.10774425e-05 | > | 3.74396503e-08 |
| 5 | -4.83348306e-07 | < | 1.27797758e-10 |
| 6 | 3.72150586e-09 | > | 1.50655827e-13 |
| 7 | -1.40903732e-11 | < | -2.49729673e-15 |
| 8 | 2.09435600e-14 | > | -3.20822166e-17 |

According to the above table, coefficients 4, 6, and 8 are reduced in the elastic net with degree 8.