|  |  |
| --- | --- |
| Laura Dooley  7/21/2024  Purpose: To build a regression model that will predict the weight based on height using the Howell.csv dataset | |
| Submission 1: Screen shot preparation code. |  |
| Submission 2: Screen shot of train/test performance. |  |
| Submission 3: Underfitting argument | A linear regression model is underfitting the training set when the model doesn’t capture the underlying patterns in the data. Indicators of an underfitting model are low R2 scores and high RMSE scores for both the testing and training set. For this linear regression model for project 4, the model is not underfitting the training. The training and testing performance for RMSE and R2 are strong and similar. The R2 values are high and the RMSE are low indicating the underlying patterns have been adequately captured in the model. |
| Submission 4: screenshot of linear fit |  |
| Submission 5: can we improve the fit of the model significantly by adding more training instances? Does this indicate under or over fitting by the model? | When the training instances were increased, the training data’s R2 was higher and the RMSE was lower than the test data. This is an indication of overfitting by the model. |
| Submission 6: Did adding the feature improve the performance of the model? Propose an explanation for the results. | When adding the extra feature, the model did not improve its performance. This outcome could be caused by the data including children. The data will include growth spurts making the graph less linear. |
| Submission 7:  Screenshot of scatter plot with cubic fit |  |
| Submission 8: Did the polynomial fit do better? Where does it fit the best? . | The polynomial did work better. There was a slight curve to the graph and the plot line was able to fit the curve. |
| Submission 9: Screenshot of scatter graph and degree 8 polynomial fit. |  |
| Submission 10: Compare the results with power=3 and power=8. Argue on whether the increase in degree is warranted by improved performance of the model | The plot line was a better fit when the plot was using the power of 3. The top portion of the graph the power=3 the plot line was more centered throughout the graph. The results for power=3 are indicating underfitting. When using power=8, the model becomes a better fit. |
| Submission 11:  Screen shot of scatter graph and degree 8 polynomial fit with Elastic net. |  |
| Submission 12: Compare the coefficients for the regular fit with power=8 to the Elastic net. Which coefficients were reduced (closer to zero). What does this tell us about our data? | The Elastic net had lower coefficients than the polynomial regression. This indicates that the Elastic net was reducing the noise in the data and regularizing the model . |