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

Tanner Young

Module 4: Lab - Building a Regression Model

A screenshot of a computer program

Description automatically generated

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

A screenshot of a computer program

Description automatically generated

Progress 3 - Analysis

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

The R^2 values for both the training and test sets are relatively high, which is a good sign. Additionally, the MAE and RMSE values on the training set are reasonable, and the values on the test set are only slightly worse. These metrics indicate that the model is making predictions that are reasonably close to the actual values on both the training and test sets.

Progress 4 - Screen shot linear fit

A screenshot of a computer

Description automatically generated

Submission 5 - Analysis

1. Will more training data improve the fit?

Yes, adding more training instances might help improve the model's performance on the validation set

b) Is the model under or over fitting?

The model is over fitting .

Submission 6 - Explain

* Did adding age improve the model?

The R^2 value on the test set has increased from 0.8677 to 0.8739. Adding the feature of age to the linear regression model seems to have improved the model's performance, at least in terms of the R^2 value.

* Propose an explanation for the results.

Age may be correlated with the target variable. If there is a relationship between age and height, age as a feature allows the model to capture this correlation and make more accurate predictions.

Progress 7 - Scatter graph

* Scatter graph with cubic model displayed
* Propose an explanation

A screen shot of a computer

Description automatically generated

Submission 8 - Explain

1. Does the polynomial fit do better?

Yes, the polynomial fit better follows the points plotted.

b) Where does it fit the best?

The polynomial fits best between a height of 140-180

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

A screenshot of a computer

Description automatically generated

Submission 10 - Explain

1. Compare degree 3 and degree 8 fits

The model with degree 8 performs slightly better on the training set in terms of both MAE and R^2. This indicates that the higher-degree polynomial model can fit the training data more closely, which is expected because it's more flexible. On the test set, both models have a similar R^2 value (0.91).

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

Yes, the increase in degree from 3 to 8 appears to be warranted as it leads to improved performance on both the training and test sets.

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

A screenshot of a computer program

Description automatically generated

Submission 12- Explain

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

Linear Regression with Power=8:

Coefficients: [-1.05821517e-06, -5.84555772e-05, -1.70146022e-03, 6.23656931e-05, -9.32819356e-07, 7.00941619e-09, -2.62048588e-11, 3.88233350e-14]

Elastic Net with Power=8:

Coefficients: [-2.55756771e-01, 6.87204298e-04, 8.52528900e-06, 3.67563933e-08, 1.17760939e-10, 1.01431642e-13, -2.55533600e-15, -3.04757085e-17]

* Which ones have been reduced in elastic net?

In the Elastic Net model, many of the coefficients are much closer to zero compared to the linear regression model. Specifically, the coefficients corresponding to higher-degree polynomial terms are reduced significantly in the Elastic Net. The coefficients for the degree 5, 6, 7, and 8 terms are notably smaller in the Elastic Net model.