

Final Paper

Kotomi Oda, Kaleb Cervantes, Nikhil Taringonda

2022-05-01

Which characteristics of Electric Vehicles have a significant impact on increasing their price?

In order to select the model, we began with a full — and untransformed — additive model. The purpose of this was to recognize which variables introduced a lot of multicollinearity. From this, the scatterplot matrix, and the correlation matrix, we were able to recognize two pairs of variables that had a lot of multicollinearity: **battery_capacity** and **Range**, **Acceleration** and **TopSpeed**. We ended up dropping the variables **battery_capacity** and **Acceleration**. This was because these predictors were less accurate in later parts of the process than **Range** and **TopSpeed** respectively.

After this, there were still possible transformations needed for the predictors. Initial predictions were found by looking at marginal plots between the predictors and the response. Other transformations of the predictors would be tested and the more accurate predictions would be kept.

There were also possible transformations needed for the response variable. This was done by using a Box-Cox Power Transformation. In this case, we got $\lambda \approx -0.5$, which corresponds to an inverse square root transformation.

This resulted in the final model:

$$\begin{aligned} \frac{1}{\sqrt{\text{PriceinUK}}} &= \beta_0 \\ &+ \beta_1 \ln \text{Range} \\ &+ \beta_2 \text{TopSpeed} \\ &+ \beta_3 \text{TopSpeed}^2 \\ &+ \beta_4 \text{Efficiency} \\ &+ \beta_5 \text{Efficiency}^2 \\ &+ \epsilon \end{aligned}$$