**Step-by-Step example for Feature Selection**

**1. Let's perform Feature Selection on the auto\_mpg dataset below.**

Table

Description automatically generated

**2. First, we compute the Pearson Correlation for the dataset.**

Table

Description automatically generated

**3. Next, look for candidates that are positively or negatively correlated with respect to our label. We are free to chose the threshold. In the example below, we only want features that are either 0.5 negatively or positively-correlated to our label (mpg, in our example).**

Table

Description automatically generated

**4. For each feature X that remains, find other features that have a positive correlation with it. Again, we can chose a threshold for this operation. In our example, we choose a threshold of 0.6.**

Table

Description automatically generated

**5. Among the features that have a positive correlation with X, and the feature X itself, select the one that has the highest correlation (either negatively or positively) with our label. The feature "weight" is most negatively correlated to our label (mpg), hence it is selected.**

Graphical user interface, application, table

Description automatically generated

**6. Discard those features that have been processed for this iteration, which are the features "cylinders" and "displacement".**

Diagram, table

Description automatically generated

**7. Now, we perform the next iteration. From the remaining features, we pick a feature X and find other remaining features that have a positive-correlation. The next remaining feature is "model year". The other remaining feature "origin" is not positively correlated to it. Hence, the feature "model year" is selected.**

Diagram

Description automatically generated

**8. We are now left with the final feature "origin".**

Diagram

Description automatically generated

**9. In our third iteration, there are no other features that are positively correlated to feature "origin". As it is the only feature (in this iteration) that has the highest correlation with the label, we accept the feature "origin" and stop.**

Diagram

Description automatically generated

**10. Our final selected features are "weight", "model year" and "origin".**

Table

Description automatically generated