Brandenburgische
Technische Universität
Cottbus - Senftenberg

Exercises

Global Model-Agnostic Interpretation Methods

Exercises

Brandenburgische
Technische Universität
Cottbus - Senftenberg

- How can you check NaN values?
- What are you doing with nominal values?
- How can you scale values?
- What are Shapley values?

Exercises

Brandenburgische
Technische Universität
Cottbus - Senftenberg

- Read data from *fueleconomy.csv*
- Use 100 samples
- Use *comb08* as target
- And as X
- Use for model: *RandomForestClassifier and*
- *shap.TreeExplainer* for Shapley Values
- Plot the Shapley summary
- Plot the feature correlation
- Plot Mean Square Error of train and test data and r2 score

Exercises

Brandenburgische
Technische Universität
Cottbus - Senftenberg

- Plot feature importance as bar chart
- Test the spearman correlation hypothesis between *co2* and *Year*
- Plot a Shapley dependence plot between *co2* and *Year*

Exercises

Brandenburgische
Technische Universität
Cottbus - Senftenberg

- What can ALE better show than Shapley values?
- Print dependency ALE plots of
  - pv2; 2-door passenger volume (in cubic feet)
  - hlv; hatchback luggage volume (in cubic feet)
- to the output *comb08*
- use the XGBoost model and package ALEPython
- What is the interpretation of the ALE plots?
-

Exercises

Brandenburgische
Technische Universität
Cottbus - Senftenberg

Global surrogates

- Create a NN with a preprocessing layer which normalizes continuous features, a layer with 64 neurons (activation='relu') and an output layer for 1 output
- Use the following optimizer and early stopping methods

```
fitted_nn_model.compile(loss='mean_squared_error',\
                optimizer=tf.keras.optimizers.Adam(lr = 0.0005),\
                metrics=['mse'])
fitted_nn_model.summary()

es = tf.keras.callbacks.EarlyStopping(monitor='val_loss', mode='min', verbose=1,\
                                      patience=200, min_delta=0.0005,\
                                          restore_best_weights=True)
```

Exercises

Brandenburgische
Technische Universität
Cottbus - Senftenberg

- Use Decision Tree and RuleFit as surrogates
- Compute RMSE for both surrogates connected with NN
- Interpret the RMSE
- Plot the Decision Tree and the important rules
- What are the most important rules?

Solutions

Brandenburgische Technische Universität Cottbus - Senftenberg

How can you explain correlation between ghgScore and co2?

Feature Interaction

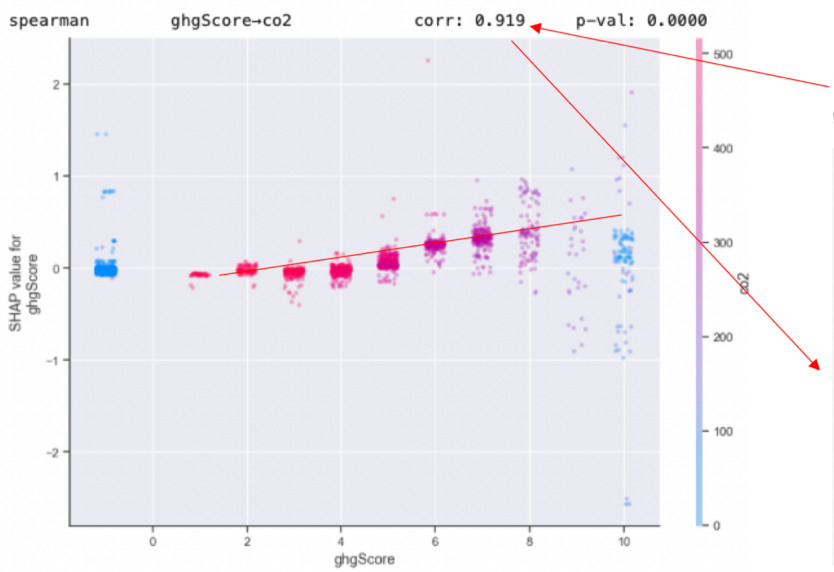
SHAP values

co2; tailpipe Co2 in grams/mile after 2013,
ghgScore; EPA (Environmental Protection Agency) score 0((worst))-10 (best), -1: not available

Brandenburgische Technische Universität Cottbus - Senftenberg

Dependence between *ghgScore and co2*

outliers



High correlation between ghgScore and co2

Greenhouse Gas Rating

MY 2024

Rating	MPG (gas)	CO_2 (g/mile)
10	>=92	0-97
9	59-91	98-152
8	43-58	153-209
7	34-42	210-265
6	28-33	266-323
5	22-27	324-413
4	18-21	414-508
3	16-17	509-573
2	14-15	574-658
1	<=13	>=659