eRum exercises

Mateusz Staniak

Exercises

Exercise 1

Fit random forest, linear regression and SVM to the housing prices data. Create DALEX explainer object for each of the models. Create and compare boxplots of residuals for all the models. Which model is the best?

Exercise 2

Draw plots of fitted vs observed values for each of these models. Can you spot any problems with the predictions? Are the prices usually underestimated of overestimated? Identify the worst predictions.

Exercise 3

Create single prediction explainers for a few of the worst predictions identified in Exercise 2. Create Break Down plots for each of the observations. What are the keys factors that drive the prediction?

Exercise 4

Use shapleyR package to calculate Shapley values for predictions examined in Exercise 3. Are the results consistent with Break Down results?

Exercise 5

Create variable importance explainer. Compare global variable importance to scores obtained in Exercise 4 and Exercise 3 for chosen observations.

Exercise 6

Simulate new data around on of the *bad* predictions from Exercise 3 using `sample_locally` function and then add predictions from black box models (`add_predictions` function). Then fit a linear model locally (`fit_explanation` function). Create waterfall plot and forest plot for the fitted model.

Solutions

Exercise 1

Exercise 2

Exercise 3

Exercise 4