DALEX CHEAT SHEET

The DALEX package (Descriptive mAchine Learning EXplanations) helps to understand how complex models are working.

Main wrapper

explain(model, data, y, predict_function, residual_function)

Function turns models into explainers - wrappers with uniform structure. Then we can use various functions to turn explainers to explanations.

model
Object - a model to be explained

data
Data.frame or matrix - data that was used for fitting.

y
Numeric vector with outputs. If provided then it shall have the same size as data.

predict_function
Function that takes two arguments: model and new data and returns numeric vector with predictions.

residual_function
Function that takes three arguments: model, data and response vector y. It should return a numeric vector.

Model understanding

model_performance(explainer)

Prepare a data frame with model residuals.

explainer
Object - a model to be explained, preprocessed by the explain function.

variable_importance(explainer, loss_function)

Calculate model agnostic variable importance.

loss_function
Function that will be used to assess variable importance.

single_variable(explainer, variable, type)

Calculates the average model response as a function of a single selected variable.

variable
character - name of a single variable

type
'pdp' for Partial Dependency and 'ale' for Accumulated Local Effects

Prediction analysis

prediction_breakdown(explainer, observation)

Calculate Break Down Explanations.

observation
A a new observarvation for which predictions need to be explained

ceteris_paribus(explainer, observations)

This function calculate ceteris paribus profiles for selected data points.

observations
set of observarvation for which profiles are to be calculated

Install the package with install_packages('DALEX')