DSSV 2023 Data Science, Statistics & Visualisation
ECDA 2023 European Conference on Data Analysis

Visualising interpretability of random forest models

PC Manefeldt[1], MMC Lamont[1], and S Lubbe[1]

[1]MuViSU (Centre for Multi-dimensional Data Visualisation), Department of Statistics and Actuarial Science, Stellenbosch University, Stellenbosch, South Africa

Models that use prediction proficiency as their aim are often viewed as black box models. So called "black box" models are able to map highly complex non-linear relationships with high order interactions, but lack interpretability. Random Forests are one such model, with decision boundaries described by its thousands of trees. Random Forests have been shown to incur a low generalisation error while also needing very little to no optimisation by the user. Random Forest Proximities and out-of-bag (oob) Random Forest Proximities act as two unique similarity similarity between observations. Multidimensional scaling (MDS) seeks to find a low dimensional approximation of pairwise similarities that provides a visual representation of the similarity between observations. By applying MDS to the Random Forest proximity measure, Random Forest proximity plots are constructed.

The Random Forest proximity plot provides a view of how the observations in your sample are related from the model's perspective, thus allowing us to "see through the eyes" of the black box. The MDS method under consideration is classical scaling, as this provides a transformation that can be used to embed new/hypothetical cases on the proximity plot. How would the model have viewed a given observation if one of its covariates were different? We can answer this counterfactual question by embedding counterfactual cases into the proximity plot.

These embedded counterfactual cases can be used to create trajectory axes. Case based trajectory axes embedded in the proximity plot, would result in a Random Forest proximity biplot, as shown below. As a special case of non-linear biplots, this enables the exploration of the relationships uncovered by the model.

Additional procedures are added to conduct inference on the sampling variability in the MDS procedure as well as diagnostics on influential cases in the MDS procedure for both the Random Forest Proximity and Random Forest oob Proximity plots.

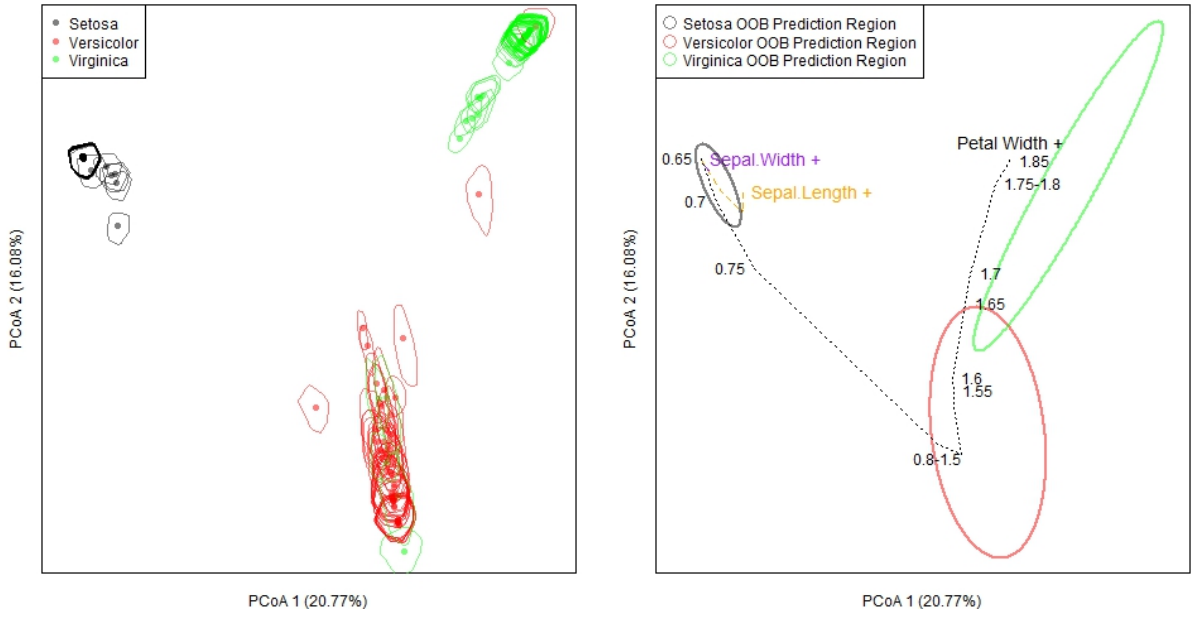

Figure 1: Left: Random Forest Proximity Plot with added 90% bootstrap alpha bags to indicate PCoA sampling variability. Right: Spanning ellipses indicating oob Prediction regions with added trajectory axes for the 1^{st} observation.

Keywords: Random Forest; Multi-dimensional Scaling; Biplot; Trajectory Axis; XAI