Exercise Sheet #6: Trees and Forests and Supports etc.

Mhd Jawad Al Rahwanji, Wojciech Ptaś

7038980, 7042843

January 24, 2023

Problem 1 – Interpretability

(a) *Interpretability* is our ability to interpret model inference. In other words, it is a measure of how easy it is for the user to understand the reasoning behind the output.

(b) We rank the methods by their interpretability thus:

1. *Decision Trees.* Tracing the decision path down the tree shows exactly how the model got to the result.

2. *Generalized Linear Models.* The intercept along with coefficient estimates can illustrate the weighting of each predictor and how it contributes to the estimation of the response.

3. *Ridge Regression.* Similar to *GLMs* but regularizes the coefficients and scales them down without removing them.

4. *LASSO.* Similar to *Ridge Regression* but uses L1-Norms so it allows predictor exclusion. Hence, it may produce more compact illustrations of predictor contribution.

5. *Random Forests.* The ensemble models produce results that cannot be traced back to a single model nor to the original predictors for that matter.

6. *Neural Networks.* Described as black boxes, are the toughest to interpret since they're made up of large weight matrices with hardly any patterns.

(c) Of course. The model *M*, in this case *Neural Networks*, would cease to be a black box. On another note, *Distillation* would effectively provide us with an equivalent *Decision Tree* that is cheaper to compute with a marginal performance trade-off.