Notes: Volatility Prediction and Uncertainty Quantification using Bayesian Neural Networks

Elias Søvik Gunnarsson

Spring, 2023

Boruta Algorithm

In some applications, variables which have no correlation to the independent variable serve as pure noise and might introduce bias in the model and reduce the predictive performance. By applying feature selection techniques we can gain some insight int the process and can improve the computation requirement and prediction accuray [**?**]. Chandrashekar_2014 classifies variable elimination methods into filer and wrapper methods, where filter methods acts as preprocessing to rank the features to be selected to be applied to a predictor. In wrapper methods predictors are wrapped on a search algorithm which finds a subset which gives the highest predictor performance.

Idea:

Let $\boldsymbol{X} \in R^{n \times m}$ be the original dataset with n samples and m features, and let $\boldsymbol{X'} \in R^{n \times m}$ be the shadow dataset obtained by randomly permuting the feature values in $\boldsymbol{X}$. Let $\boldsymbol{y} \in R^n$ be the target variable.

1. $\boldsymbol{X'} = permute(\boldsymbol{X})$
2. Train a random forest algorithm on $\boldsymbol{X}$ and calculate the feature importance scores, imp_i, for each feature i.
3. For each feature i, compare its importance score imp_i to the maximum importance score imp'_i in the shadow dataset. If $\text{imp}_i \geq \text{imp}'_i$, mark feature i as confirmed and add it to the final feature set $\boldsymbol{F}$. Otherwise, mark feature i as tentative and keep it for further evaluation.
4. Repeat steps 1-3 until all features have been either confirmed or rejected. At each iteration, use the confirmed features $\boldsymbol{F}$ as input to train the random forest algorithm.
5. For each tentative feature i, perform the shadow feature test by randomly permuting the feature values in $\boldsymbol{X'}$ and calculating the feature importance scores, imp'_i. If $\text{imp}_i \geq \max(\text{imp}'_i)$, mark feature i as confirmed and add it to $\boldsymbol{F}$. Otherwise, mark feature i as rejected and remove it from consideration.
6. Return the final set of confirmed features $\boldsymbol{F}$.