INFTrees and INFforests Variable Importance

Theory

The theory behind INFtrees combines the permuatation approach to variable importance found in Strobl et al with Breiman et al 1984's notion of grouped predictors.

Grouped Predictors

INFTrees

For a CART, T, representing the model $Y\ X_1, ..., X_p$ where $Y, X_1, ..., X_p$ are vectors of length n, the INFTrees algorithm proceeds as follows:

Algorithm 1 INFTree, $VI_{inf}(T)$

for each $X_i \in X_1, ..., X_p$ **do**
  Calculate: $\Phi_o = RSS(T, (Y, X_1, ..X_p))$
  Fit the tree T_{X_i}, where $T_{X_i} : X_i \sim X_1, ..., X_{i-1}, X_{i+1}, ...X_p$
  Extract the set P_{X_i} of partitions on X_i from T_{X_i}
  Permute X_i with respect to P_{X_i}
  Find $\Phi^* = RSS(T, (Y, X_1, ..., \bar{X}_i, ...X_p))$
  The difference between these values, $\Phi^* - \Phi_o$, is the variable importance for X_j on T,
end for

This procedure allows the null hypothesis that Y is independent of X_i given the values of $X_1, ...X_{i-1}, X_{i+1}, ...X_p$ to be tested. Therefor, values of VI_{inf} could be compared in a similar manner to the coefficients of linear regression.

INFForests

The algorithm for determining $VI_{inf}(R)$ follows similarly.

Algorithm 2 INFForests, $VI_{inf}(R)$

1: Fit a random forest, R on the dataset D fitting the model $Y \sim X_1, ..., X_p$.
2: **for** each $X_i \in X_1, ..., X_p$ **do**
3: **for** each $t \in R$ **do**
4: Calculate: $\Xi_o = \frac{1}{\nu_t} RSS(t, \bar{B}^t)$
5: Calculate a tree T_i that predicts $X_i \sim X_1, ..., X_{i-1}, X_{i+1}, ...X_p$ using the subset of the observations used to fit t
6: Permute the subset of X_i contained in $\bar{B}_t$ with respect to the set of partions P_{xi} from T_i.
7: Now find $\Xi^* = \frac{1}{\nu_t} RSS(t, \bar{B}_t^*)$
8: The difference between these values, $\Xi^* - \Xi_o$, is the variable importance for X_i on t
9: **end for**
10: Average over all $t \in R$

$$VI_{inf}(X_i, R) = \frac{1}{ntree} \sum^{ntree} \Xi^* - \Xi_o$$

$$VI_{inf}(X_j, R) = \frac{1}{ntree} \sum^{ntree} \frac{1}{\nu_t} RSS(t, \bar{B}_t^*) - \frac{1}{\nu_t} RSS(t, \bar{B}^t)$$

11: **end for**