1 Functional modules of EnsembleFS R package

Table 1 presents the basic functional modules of the EnsembleFS R package.

Table 1: Overview of functional modules in EnsembleFS R package. Note, see text.

Module	Description
fs.utest, fs.mcfs, fs.mdfs.1D, fs.mdfs.2D, fs.mrmr	fs.utest(), fs.mcfs(), fs.mdfs.1D(), fs.mdfs.2D(), and fs.mrmr() functions perform feature selection (FS) on data using feature filter: U-test, MCFS, MDFS-2D, MDFS-1D, and MRMR, respectively.
cross.validation	cross.val() run the resampling procedure, namely, the stratified k-fold cross-validation or the stratified random sampling n-times on data.
feature.selection	feature.selection() perform ensemble FS on the whole dataset; feature.selection.cv() perform ensemble FS within resampling loop.
correlation	correlation.search(), correlation.removed() search and remove the correlated variables with the training set, respectively.
rf.model	build.model() built individual random forest models on data and compute performance metrics; build.model.crossval() built random forest models within the resampling loop.
stability.selection	compute the Lustgarten's stability measure ASM for top N feature across all models.
features.ranking	ranking.feature() computes the ranking of most informative features based on the frequency of feature occurrences across all models.
graph.result	plot the graphs of the mean values of $ASM(N)$, $AUC(N)$, $ACC(N)$, and $MCC(N)$ for top N = 5, 10, 15, 20, ..., 50, 75, 100 features.
gene.information	get.info.gprofiler() get biological information about biomarkers from the Gene Ontology, the Reactome, the WikiPathways, the miRTarBase, the Kyoto Encyclopedia of Genes and Genomes, the Human Protein Atlas, the Transfac, the CORUM, and the Human Phenotype Ontology databases; get.info.top.gene() get information about the biomarkers that are selected using the ensemble approach.
main	ensembleFS() perform ensemble feature selection on training data sets and build the random forest classification models for evaluation of feature selection algorithms in n-resampling steps.

References