

# Description of aFANS-related Functions

## **L50\_sta(sample1, sample2)**

Description: *Calculate the L50 statistic of sample1 and sample2.*

Input: *Two numeric vectors, sample1 and sample2.*

## **PL50\_var\_judge(dataset, yname="species", y=c(1,0), alpha = 0.05, iter = 1000)**

Description: *Test the heteroscedasticity of dataset.*

Input:

- *dataset: Dataset for test.*
- *yname: The name of label.*
- *y: Value of label.*
- *alpha: The  $\alpha$  for test.*
- *iter: Iteration times.*

Output: *A label list of homoscedastic and heteroscedastic feature.*

## **lr\_rate(dataset, iter = 200, subtrain=150, subtest=300)**

Description: *Test the efficiency of LR model.*

Input:

- *dataset: Data pool for test.*
- *iter: Iteration times.*
- *subtrain: Number of training set.*
- *subtest: Number of testing set.*

Output:

- *rate: The accuracy, false positive rate, false negative rate, and AUC.*
- *pro\_label: The pair of probability and label for plotting ROC curve.*
- *time: The execution times.*

**fans\_lr\_rate(dataset, iter = 200, subtrain=150, subtest=300, L=10)**

Description: *Test the efficiency of FANS model.*

Input:

- *dataset: Data pool for test.*
- *iter: Iteration times.*
- *subtrain: Number of training set.  $2 * \text{subtrain} + \text{subtest}$  should less than  $\text{nrow}(\text{dataset})$ , because some data are used for density estimation.*
- *subtest: Nnumber of testing set.*
- *L: L in the algorithm.*

Output:

- *rate: The accuracy, false positive rate, false negative rate, and AUC.*
- *pro\_label: The pair of probability and label for plotting ROC curve.*
- *time: The execution times.*

**fans2\_lr\_rate(dataset, iter = 200, subtrain=150, subtest=300, L=10)**

Description: *Test the efficiency of FANS2 model.*

Input:

- *dataset: Data pool for test.*
- *iter: Iteration times.*
- *subtrain: Number of training set.  $2 * \text{subtrain} + \text{subtest}$  should less than  $\text{nrow}(\text{dataset})$ , because some data are used for density estimation.*
- *subtest: Number of testing set.*
- *L: L in the algorithm.*

Output:

- *rate: The accuracy, false positive rate, false negative rate, and AUC.*
- *pro\_label: The pair of probability and label for plotting ROC curve.*
- *time: The execution times.*

`afans_lr_rate(dataset, iter = 200, subtrain=150, subtest=300, L=10, L=10, alpha_var=0.05, var_iter=1000)`

Description: *Test the efficiency of aFANS model.*

Input:

- *dataset*: Data pool for test.
- *iter*: Iteration times.
- *subtrain*: Number of training set.  $2 * \text{subtrain} + \text{subtest}$  should less than  $\text{nrow}(\text{dataset})$ , because some data are used for density estimation.
- *subtest*: Number of testing set.
- *L*:  $L$  in the algorithm.
- *alpha\_var*:  $\alpha$  for PL50 test.
- *var\_iter*: Iteration times for PL50 test.

Output:

- *rate*: The accuracy, false positive rate, false negative rate, and AUC.
- *pro\_label*: The pair of probability and label for plotting ROC curve.
- *time*: The execution times.