binspec

May 26, 2016

`binary_peaks` *Find binary peaks*

Description

Find peaks in window of size 2*neighbors + 1 and label m/z integers within the error as peaks. Returns vector of peak m/z integers.

Usage

```
binary_peaks(df, neighbors, error = 0)
```

Arguments

`df`	Data frame of m/z and intensities
`neighbors`	Number of neighboring m/z values to compare on right and left
`error`	m/z Decimal error value

`classifier_accuracies`
Classifier Accuracies

Description

Find the best classifier using leave-one-out cross validation (svm) and out-of-bag error (random forests). Returns a list of classifier results

Usage

```
classifier_accuracies(peaks, labels, min_peak_percentage)
```

Arguments

`peaks`	Boolean matrix of mass spectra rows with m/z columns, indicating if an m/z value corresponds to a peak.
`labels`	The correct classifications of the peaks.
`minpeaks`	How many "true" values must show up for a given m/z value for it to be considered a feature.

`combine_peaks` *Combine peak vectors*

Description

Create a binary matrix, each column represents an m/z value, and each row represents a mass spectra. The value indicates whether or not the m/z of this spectra is a peak.

Usage

```
combine_peaks(list_mz_peaks)
```

Arguments

`list_mz_peaks`	List of m/z peak vectors

`naive_feature_importance` *Rank importance of features*

Description

Given a matrix of binary peaks and each row's corresponding labels, this function takes returns the absolute difference between the proportion of times an m/z value was labeled as a peak within each of the two classes.

Usage

```
naive_feature_importance(peaks, labels)
```

Arguments

`peaks`	A matrix of peaks
`labels`	A factor vector of labels whose length is equal to the number of rows of peaks

`round_df` *Round data frame*

Description

Round all m/z and intensity values to integers.

Usage

```
round_df(df)
```

Arguments

`df`	Data frame

svm_rf *SVM and RF Accuracies*

Description

Given a vector of neighbor values and a vector of the minimum number of peaks to be considered, this function finds the peak mz values for a data set by running binary_peaks using each of the neighbor vector values, runs SVM and RF on the peaks for each of the min_peak_count values, and returns the accuracies of each test in a table. The table's rows are the number of neighbors, and the columns are the min_peak_count values.

Usage

```
svm_rf(list_of_dfs, labels, neighbors, min_peaks_percentage)
```

Arguments

list_of_dfs	The first data frame of mz values and frequencies
labels	The labels of the two states the first data frame's values could be classified as
neighbors	A vector of the number of neighbors to be considered in the binary_peaks function
min_peaks_count	A vector of the minimum numbers of peaks to be considered in the classifier_accuracies function

Index