binspec

June 10, 2016

`accuracy_matrix` *Generates matrix of accuracy values*

Description

Returns a matrix of accuracy values where the columns represent the classifier type, and the rows represent the min_peak_count.neighbor values used for the classifier

Usage

```
accuracy_matrix(rf, svm)
```

Arguments

`rf`	A random forest object
`svm`	A support vector machine object

`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

peak_frequencies *Get frequency of peaks*

Description

Return frequency of peaks within each label

Usage

```
peak_frequencies(peaks, labels)
```

plot_naive_importance *Plot importance of naive importance vector*

Description

Plot importance of each m/z value according to the naive ranking vector. Returns a ggplot object

Usage

```
plot_naive_importance(naive_importance, count = 10)
```

Arguments

naive_importance	A vector returned from naive_feature_importance()

plot_rf_differences *Plot important peaks on each day*

Description

Returns a ggplot object of the most important peak frequencies

Usage

```
plot_rf_differences(peak_freqs, rf, count)
```

Arguments

rf	A random forest object, to be used for feature importance
count	Number of peaks to select
peaks_freqs	The frequency of each peak within each label

plot_rf_importance *Plot importance of randomForest*

Description

Plot importance of each m/z value according to a randomForest object. Returns a ggplot object

Usage

```
plot_rf_importance(rf, count)
```

Arguments

rf	A randomForest object

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,
  error_window = 0.005)
```

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_percentage	A vector of the minimum percent of times an m/z must be a peak to be considered in the classifier_accuracies function
errow_window	A vector of percentage of nearby peaks that should be also labeled as peaks when one is found

Index