

MWAS code structure:

`mwas_analysis.r` -- command line function, includes "train", "predict", "plot", "statistics" options.

Dependencies (required packages):

1. `optparse` - option parsing
2. `biom` - biom format file operation
3. `e1071` - classifier package (SVM, cross-validation...)
4. `randomForest` - random forest
5. `glmnet` - for multi-nomial regression
6. `pROC` - ROC analysis
7. `caret` - variable importance analysis for feature selection
8. `xlsx` - for Excel workbook output
9. `beeswarm` - beeswarm plot
- # 10. `kernlab` - future version will add this function

Input Option list

Training:

```
Rscript mwas_analysis.r -w train
-i   otu_fp
-m   map_fp
-c   category
-o   out_dir
-t   classifier_method
-k   classifier_parameter
-e   validation_type
-f   number_of_fold_cv
-s   is_feature_selection
-b   feature_selection_paramter
```

```
Rscript bin/mwas_analysis.R -w train -i
test/data/GG_100nt_even10k-adults-s20.biom -m test/data/gg-
map-adults.txt -c COUNTRY -o example/train_out/ -t svm
```

Predicting:

```
Rscript mwas_analysis.r -w predict
-i   otu_fp
-m   map_fp (optional)
-c   category (optional)
-o   out_dir
-k   trained_model_fp
```

```
Rscript bin/mwas_analysis.R -w predict -i
test/data/GG_100nt_even10k-adults-s20.biom -m test/data/gg-
map-adults.txt -c COUNTRY -o example/predict_out/ -k
example/train_out/trained_model.rds
```

Plot: (all capital for plot-only parameters)

```
Rscript mwas_analysis.r -w plot
```

```
-i      otu_fp
-o      out_dir
-V      plot_type
-F      feat_fp
-D      distance_fp
-P      pcoa_fp
-W      which_taxa
-S      shorten_taxa
-M      multiple_axis
-N      nplot
-T      transform_type
-X      x_axis_label
-C      suppress_relative_abundance_conversion
-R      sort_by_abundance
-A      Maximum false discovery rate (alpha)
-O      category_order
```

```
Rscript bin/mwas_analysis.R -w plot -V gradients -i
test/data/GG_100nt_even10k-adults-s20.biom -W
'Bacteroides,Prevotella' -o example/plot_out_gradient
```

```
Rscript bin/mwas_analysis.R -w plot -V diff -i
test/data/GG_100nt_even10k-adults-s20.biom -W
'Bacteroides,Prevotella' -o example/plot_ou_diff
```

Statistics:

```
Rscript mwas_analysis.r -w statistics
```

```
-i      otu_fp
-m      map_fp
-c      category
-o      out_dir
```

Function list

File name	Function names	Notes
import_data.R	import.train.params	
	import.predict.params	
	import.plot.params	
	import.stats.params	
util.load.R	read.qiime.table.mwas	
	read.qiime.classic.table	
	load.qiime.mapping.file	
	load.qiime.otu.table	
	otu.table.has.metadata	
	get.header.index	
	load.qiime.taxon.table	
	load.qiime.distance.matrix	
	load.qiime.pcoa.file	
	load.qiime.feature.stats	
	remove.nonoverlapping.samples	
	load.experiment	
	cluster.by.correlation	
	collapse.by.correlation	
	inspect.env	
	shorten.taxonomy	
	balanced.folds	
	is.outlier	
	factor.to.numeric	
	parse.params	not used
model.train.R	train.mwas	
	persist.model.mwas	
	cross.validation.mwas	
model.evaluation.R	model.evaluation.mwas	
	predict.knn	S3 method
preprocess_mwas.R	preprocess.mwas	
	filter.pathways	
	get.next.kegg	
visualization.R	plot.mwas	
	plot.differentiated.taxa	
	plot.beeswarm.dt	
	run.scatterplot	not used
	run.scatterplot.all.data	not used
	run.beeswarm	not used
	plot.gradients	
	qqplot.pvals	not used

export_data.R	export.mwas	
	save.xlsx	
	write.differentiation.test.results	
stats.r	differentiation.test	
	linear.test	not used
	t.test.wrapper	not used
roc_mwas.R	roc.mwas	
gradients.R	show.gradients	
	show.metadata	
	get.gradient.ixs	
	best.legend.location	
features.selection.R	feature.scores.rf.mwas	
heatmap.3.R	heatmap.3	
heatmap.mwas.R	cluster.columns	
	cluster.rows	
	shorten.taxonomy	Similar to the one in util.load.R; commented out
	create.color.bars	
	heatmap.mwas	