READ ME

**##############################################################################**

This folder contains example data and R code for selecting and combining markers in treatment selection and for estimating total disease and treatment cost of corresponding treatment-selection rules, as proposed in the paper “Identifying Optimal Biomarker Combinations for Treatment Selection through Randomized Controlled Trials” by Ying Huang.

A relevant paper is “Identifying Optimal Biomarker Combinations for Treatment Selection via a Robust Kernel Method”, *Biometrics*, 2014, 70(4): 891-901.

If you have questions, please contact Ying Huang (yhuang@fhcrc.org).

1. Contents of the folder:

1. example.Rdata: contains an example training data (data.train) used for deriving marker-based treatment selection rule, and an example test data (data.test) for estimating the total disease and treatment cost associated with a particular treatment-selection rule
2. exampleFun.R: an R script including functions needed for generating weighted-SRAMP-L1 estimator based on a training data; for generating predicted score for making treatment recommendation for new observations; and for estimating total disease and treatment cost associated with the treatment-selection rule in a test data
3. example.R: an R script demonstrating how to apply the functions included in exampleFun.R

2. Setting:

* Training / testing data are samples from a 1:1 randomized trial, TT=1,0 stands for treating and not treating respectively
* Disease outcome y is an unfavorable outcome (can be either binary or continuous).
* Biomarkers **X** (can be univariate or multivariate) are used to construct a treatment -selection rule A(**X**)=I(f(**X**)>0) with f(**X**) being some function of x. One will be recommended to receive the treatment if A(**X**)=1 and not receive treatment if A(X)=0.
* δ is the treatment-to-disease burden ratio, i.e. burden associated with treatment relative to the burden associated with one unit of Y. For binary Y, this will be the burden associated with treatment relative to the burden associated with per event of Y.
* Total burden, i.e., sum of disease burden and treatment burden =

= E(Y|A(X)=1, T=1)P(A(X)=1|T=1)+E(Y|A(X)=0,T=0)P(A(X)=0|T=0)+E(A(X))× δ

3.R library to be installed:

To run this code, you must install following R libraries available from CRAN:

* kyotil: for matrix multiplication and kernel function generation

4. Variables in the dataset

a). data.train

* yy: disease outcome
* xi: for marker i (where i=1,…, M) for M a positive integer
* TT: treatment indicator: 0 for untreated, 1 for treated
* weight: individual sampling weight

b). data.test

* yy: disease outcome
* xi: marker i
* TT: treatment indicator
* weight: individual sampling weight

5. Algorithm demonstrated in the example code:

WSRAMP-L1 estimator for minimizing the penalized weighted sum of ramp loss