MCIAbase_model

Source: Example models first challenge (CMI-PB/mcia-model)

Very good results.

Uses Multiple Co-Inertia Analysis (MCIA). Co-Inertia analysis is a multivariate method that identifies trends or co-relationships in 2 datasets. It has the benefit of being applicable to datasets where the number of variables exceeds the number of observations. CIA is accomplished by finding successive orthogonal axes with maximal squared covariance from the two datasets. (This can be done via PCA or other methods such as correspondence analysis, which is better for microarray data.) CIA can analyse multiple sets of qualitative and quantitative data. Based on an optimization criterion CIA projects 2 datasets onto the same dimensional space and transforms their features onto the same scale. MCIA then extends the technique to 3+ datasets.

Baseline_model

Source: Example models first challenge (CMI-PB/mcia-model)

Also uses MCIA

Study_2_CMI-PB_workflow

Source: Example models first challenge (literature_models_first_challenge/Study-2-Tsang-2014/Study_2_CMI-PB_workflow.Rmd)

Uses diagonal linear discriminant analysis. The DLDA is a variant of LDA, where the off-diagonal elements of the pooled sample covariance are treated as 0.The features are assumed to be uncorrelated

Study_4_Fourati_2016

Source: example models first challenge (literature_models_first_challenge/Study_4_Fourati_2016

This implementation is based on a study from Fourati et al 2015. (PMID 26742691) It utilizes 5 approaches from that study. The first was "BioAge" an aggregate score based on the difference in expression between age-related gene modules. They also used 2 versions with fewer gene modules. Lastly they used a Naive bayes classifier on a curated set of genes and a multivariate logistic regression with cytometric flow markers.

Study_7_Fourati_2021

Source: example models first challenge (literature_models_first_challenge/Study_7_Fourati_2021)

This model is a Random Forest classifier using 500 genes identified as ideal in the original paper that were also in the CMI-PB dataset.

Nixstix&CMI/PB2

Source: example modells second challenge (nixstix/CMI-PB2/scripts)

This approach uses Multi-Omics Factor Analysis (MOFA) and Lasso Regression to make predictions. MOFA is a method for finding the main sources of variation in data sets. It does this by determining a set of factors that capture the sources of variability.

Jgygi/CMIPB2_Spear

Source: example models second challenges (jgygi/CMIPB2_SPEAR/CMIPB_2nd_code_submission.Rmd)

This model uses SPEAR (Sparse Supervised Bayesian Factor Model for Multi-omics), Mofa and Lasso regression to make predictions.

Block Random Forest Model

Source: example models second challenge (nehera/CMIPB_challenge/BlockRandomForest_EFL.r)

The block forest is a random forest for blocks of data. In this a "Block" is a subset of the multi-omics data of a patient containing the individual data types. The Block forest functions by: 1) randomly obtaining a subset of all blocks, 2) sampling a fixed number of variables from each selected block, and then 3) building a decision tree on the thus chosen variables.