

Files

- twoHoldOutExp (m file)
- readCustomFile (m file)
- dataStorage (m file)
- generateInteraction (m file)
- generateGeneCpd (m file)
- geneTRCMPLXstats (m file)
- plotAUC (m file)
- geneExpression (mat file)
- probabilities (mat file)
- aucANDpredictions sample TRCMPLX.mat (mat file)

Description of function and script files (.m files)

- twoHoldOutExp

conducts the two hold out experiment and is the main file/function to start with. Takes eviDence ('ge' for gene expression, 'me' for methylation) and model ('t1' for PBK+EI, 't2' for PBK and 'p1' for NB+MPBK). Here, PBK+EI - Prior Biological Knowledge and Epigenetic Information, PBK - Prior Biological Knowledge, NB+MPBK - Naive Bayes with Minimal Prior Biological Knowledge

- readCustomFile

reads the gene expression from the geneExpression mat file and returns a set of unique genes in uniqueGenes, the gene expression matrix in expressionMatrix, the total number of genes involved noGenes, the total number of samples involved noSamples, the ground truth labels of the samples in groundTruthLabels and transGroundTruthLabels

- dataStorage

take the type of model involved stored in model ('t1'/'t2'/'p1') and stores the required probability values in workspace as well as probabilities mat file.

- generateInteraction

take the unique genes stored in uniqueGenes as well as the model ('t1'/'t2'/'p1') and generates the interaction (i.e causal arcs) among the nodes in interaction and the list of names of the nodes in nodeName. These will be used in the generation of the directed acyclic graph of the Bayesian Network.

- generateGenecpd

takes in the expression data of a particular gene for training in vecTraining, the labels pertaining to the particular gene expression data inlabelTraining, the name of the gene node in nodeName, the parents of the gene node

in parent, the probability tables of the parents of the gene node in parent_cpd and finally the model type in model ('t1'/'t2'/'p1') to return the conditional probability table for the gene node under consideration.

- geneTRCMPLXstats

This file generates the tables which shows how TRCMPLX behaves as the evidences of genes vary in both normal and tumorous cases and are depicted in the results sections of Sinha:2014. Interpretations of the results can be studied in more depth from Sinha:2014. Once the results have been saved in Results.mat file, one can rename the file based on the model and the evidence arguments used in function twoHoldOutExp. An example of one such file is Results-T1-GE-pforTRCMPLX-90per.mat which is loaded using the load command and further processed using the script in the m file titled geneTRCMPLXstats.

- plotAUC

Note that to generate the ROC graphs and their respective AUC values for different models with varying effect of TRCMPLX on different genes (ETGN in Sinha:2014), the results in variable s X and Y (of twoHoldOutExp) are stored in different variables and clumped together in a mat file titled aucANDpredictions_sample_TRCMPLX.mat. This has to be done manually for each model and every setting of ETGN. For example, using model t1 and ETGN of 60%, the false positive rate in X is stored as xT1_50 and the true positive rate in Y is stored as yT1_50, in the above mentioned .mat file. Finally, the script in the m file titled plotAUC is used to manipulate the aforementioned transformed variables and generate the ROC curves in the results section of Sinha:2014.

Files needed to run

Before running the scripts

1. one must have Matlab running on the machine
2. download and install the Bayesian Network Toolbox by Murphy.

Citations

-
- (1) If using geneExpression.mat file please cite the following - Jiang X, Tan J, Li J, Kivimäe S, Yang X, Zhuang L, Lee P, Chan M, Stanton L, Liu E, Cheyette B, Yu Q (2008). "DACT3 is an epigenetic regulator of Wnt/ β -catenin signaling in colorectal cancer and is a therapeutic target of histone modifications." Cancer cell, 13(6), 529-541.

- (2) If using figure wnt-img.pdf please cite the following - Verhaegh W, Hatzis P, Clevers H, van de Stolpe A (2011). "Personalized cancer treatment selection using computational signaling pathway models." Cancer Research, San Antonio Breast Cancer Symposium, 71(24), 524-525.
- (3) If using any other material from this work, please cite - Sinha S (2014). "Integration of prior biological knowledge and epigenetic information enhances the prediction accuracy of the Bayesian Wnt pathway." Integr. Biol., 6, 1034-1048. doi: 10.1039/c4ib00124a. AND Sinha S (2015). "A Pedagogical Walkthrough of Computational Modeling and Simulation of Wnt Signaling Pathway Using Static Causal Models in MATLAB." Unpublished Manuscript, Biorxiv.