

CyNetSVM User Manual

(A Cytoscape app for cancer biomarker identification using network constrained support vector machines)

1. Installation

1.1 Package Requirements

The package requires Java 8 and Cytoscape 3 (<http://www.cytoscape.org/>);
Tested on: Multiple Operating Systems (Windows 7, Ubuntu 12.04, Mac OS) with Java 8 and Cytoscape 3.

1.2 Install from Cytoscape App store

The software package is available at Cytoscape App store;
From menu click App -> App Manage -> Search for CyNetSVM -> Click and install.

1.3 Install from Package

The package, the java executable file for Cytoscape 3, can be downloaded at <http://sourceforge.net/projects/netsvmjava/files/CyNetSVM-v1.0.jar/download>.

To install:

- Download the jar file;
- Open Cytoscape; from menu click App -> App Manager and click 'install from file' and find CyNetSVM-v1.0.jar and click Open.

A new tab 'CyNetSVM' should appear and the GUI of the app is shown in Fig. 1.

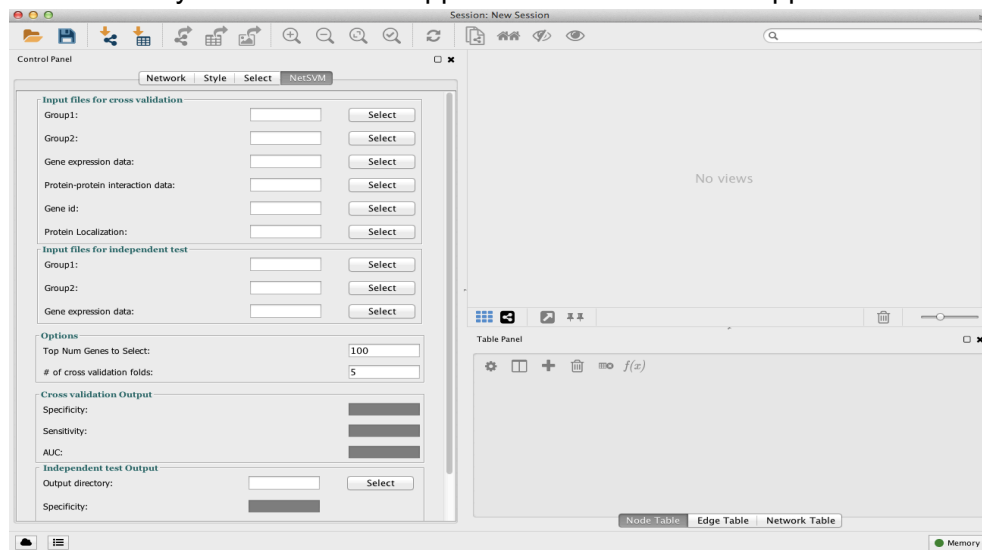


Fig. 1. A screenshot of the GUI of CyNetSVM (Cytoscape App).

2. Testing CyNetSVM with example data

The example data are made available at

<http://sourceforge.net/projects/netsvmjava/files/Example.zip/download>.

2.1 Open Cytoscape

Open Cytoscape and click on the “CyNetSVM” tab. Fig. 1 shows the GUI of the app.

Users need to provide the following input data:

Data	Format	Description
Cross- validation		
Group 1	TAB	Sample IDs of Group1 in the data
Group 2	TAB	Sample IDs of Group2 in the data
Gene expression data	GCT	Gene expression data
Protein-protein interaction data	TAB	Protein interaction networks
Gene id	TAB	Gene list of interest
Protein localization	TAB	This file contains gene IDs and their matching gene symbols, locations in the cell and participated pathways
Independent test		
Group 1	TAB	Sample IDs of Group1 in the data
Group 2	TAB	Sample IDs of Group2 in the data
Gene expression data	GCT	Gene expression data

To test on the example data, users can load the files from the “Example” folder. “EarlyIndex.txt” is a tab-delimited text file with the sample IDs of Group 1; “LateIndex.txt” is a tab-delimited text file containing the sample IDs of Group 2; “PPI.txt” is a tab-delimited text file with two columns containing the gene IDs of the protein interaction data; “Genelid.txt” is a tab-delimited text file containing the gene IDs; “GeneExp.txt” is a gene expression file in GCT format (http://www.broadinstitute.org/cancer/software/gsea/wiki/index.php/Data_formats); “entrezidmap.txt” is a tab-delimited text file containing the gene IDs mapped to their gene symbols, locations in the cell and participated pathways (separated by comma). (Note that if the independent test dataset is not provided, CyNetSVM will perform “independent test” on the whole train data to generate the results.)

In addition, the following options are set by default (which users can change as required):

Options	Description
# of nodes selected	# of nodes used to generate the final network
# of cross-validation folds	# of folds used for cross-validation used in CyNetSVM to calculate the parameters

With the input data specified, users need to set a path for the output directory (which will be used to store the predicted gene network, the top genes involved, the gene weight matched to the input gene ID and the sample predicted values in the network by

CyNetSVM). Users can click the “Start netSVM” button to start the analysis. A message will be displayed with the current status of the program (e.g., ‘creating network’).

3. Interpretation of results

3.1 Output file description

There are five output files:

- 1) “topgene.txt”: a tab-delimited text file containing the list of top ranked genes (predicted by CyNetSVM) differentiating the two conditions in the data.
- 2) “network.txt”: a tab-delimited text file containing the predicted network structure formed by the top ranked genes.
- 3) “gene_weight.txt”: The weights of genes estimated by CyNetSVM using cross-validation, which is ordered by the ascend sorting of input gene IDs.
- 4) “gene_weight_inde.txt”: The weights of genes estimated by CyNetSVM using all samples in the training data.
- 5) “raw_values.txt”: The absolute predicted values of samples (a positive number indicates the sample being predicted to Group 1 and a negative number being predicted to Group 2).

3.2 Visualize network in Cytoscape

Users need to make sure to set the correct output directory path where the predicted network structure will be written in a tab-delimited text file. This file will be later used to create and view the network in Cytoscape. Other attributes of nodes and the network (such as fold change, the contribution of each gene of the weight vector, location within the cell, specificity, sensitivity and area under the curve (AUC)) are also displayed. Fig. 2 shows a screenshot of the output from the example data. Node color indicates fold change. A node in red color represents that the node is over-expressed in Group 1; a node in green color indicates that the node is over-expressed in Group 2.

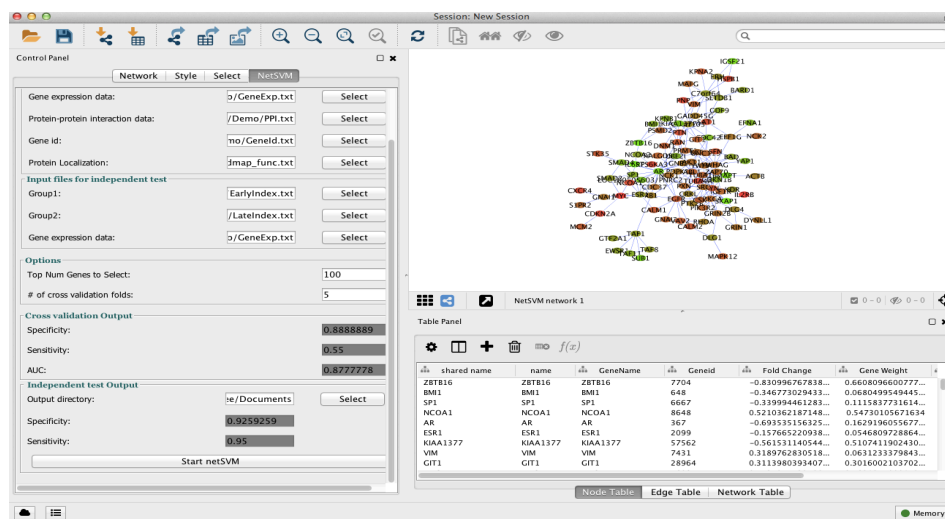


Fig. 2. A screenshot of the output from the example data.