This is source code for inferring causality by the CVP algorithm in python. The CVP is an algorithm for causal inference, the full name of which is "Causality inference based on cross-validation predictability ".

The 'CVP\_inferring\_causality.py' can be used to build the causal network by the CVP algorithm, and details of the parameters are shown below.

**Dependencies**

python 3.7 or later

numpy 1.2 or later

pingouin 0.3.10 or later

scipy 1.6.1 or later

pandas 1.0.3 or later

scikit-learn 1.0.2 or later

**Usage**

**Run the main program**

python CVP\_inferring\_causality.py -m=m -infile=infile\_name -threshold=thres -outfile=outfile\_name

-m: The parameter 'm' indicates the grouping for m-fold cross-validation. The default value is 3.

-threshold: The parameter 'threshold' is a correlation threshold produced by program 'pcc\_threshold.py', the fault threshold value is 0.

-infile: The parameter 'infile' indicates the file name for input. The input file is a data matrix that each row represents a sample and each column represents a variable. The file 'data\_example.txt' is an example file for input file.

-outfile: The parameter 'outfile' indicates the name of output file to store the causality with format '.txt'. The default value is 'CVP\_result.txt'. The output file has three columns, the first column indicates the cause variable, the second column indicates the effect variable and the third column indicates the causal strength.

**Calculate the correlation threshold for an expression data**

python pcc\_threshold.py -infile=data\_ example.txt

-infile: The parameter 'infile' indicates the file name for input. The input file is a data matrix that each row represents a sample and each column represents a variable. The file 'data\_ example.txt' is an example file for input file.

**Example 1: Using the CVP program to calculate the causality between variables**

This file 'data\_example.txt' is an example to infer causality for 9 variables with 8 samples. Each column in this file represents a variable, and each row in this file represents a sample. We can infer the causality of the 9 variables by executing below:

python CVP\_inferring\_causality.py -infile=data\_example.txt -threshold=0 -m=3

This 'CVP\_inferring\_causality.py' directly uses the default threshold for correlation as long as the number of variables is less than 1000, otherwise the corresponding correlation threshold needs to be calculated by the 'pcc\_threshold.py'.

Using the 'CVP\_inferring\_causality.py', a causal network of nine variables is constructed and the output file is the default 'CVP\_result.txt'. The output file has three columns, the first column indicates the cause variable, the second column indicates the effect variable and the third column indicates the causal strength. Each row is a causality between two variables in the 9 test variables in 'data\_example.txt'.

**Example 2: Calculate the correlation threshold**

python pcc\_threshold.py -infile=data\_example.txt

This file 'data\_ example.txt' is matrix data, with columns representing variables and rows representing samples.