1 How to use MINE

MINE is written in Java can be downloaded as a JAR from exploredata.net. The only mandatory parameters are the name of the file containing the data and a specification of which variable pairs to analyze. It is invoked as follows:

```
java -jar MINE.java   infile   masterVariable
```

The mandatory parameters may be set as follows

- infile : A path to a comma-separated values (csv) file containing the data. The variable names can either be in the first line of the file (making each row a record), or the first column in the file (making each column an entry).
- masterVariable : Set this to '-allPairs' to compare all pairs of variables against each other or to '-adjacentPairs' to compare consecutive pairs of variables only. Set to a number i to compare all variables only against the i-th variable.

In addition, the following optional parameters/flags are provided

- cv : A floating point number indicating which percentage of the records need to have data in them for both variables before those two variables are compared. Default value is 0.
- exp : The exponent in the equation $B(n) = n^{\alpha}$. Default value is 0.6.
- c : Determines by what factor clumps may outnumber columns when OptimizeXAxis is called. When trying to partition the x-axis into x columns, the algorithm will start with at most cx clumps. Default value is 15.
- gc : The number of variable pairs to analyze before forcing a Java garbage collection. This should not be necessary unless sample size is very small and there are very many variable pairs. Default value is Integer.MAX_VALUE.
- -permute : Instructs MINE to permute the dataset before running it. If masterVariable is set to '-adjacentPairs', then every other variable will be permuted. If it is set to a variable id, all variables except for the master variable will be permuted. If it is set to '-allPairs', every variable will be permuted independently of the others. This flag is unset by default.
- jobID : A string to identify this job. The program will produce two files; one is called [infile] ,[jobID],Results.csv, and the other is called [infile],[jobID],Status.txt (always contains the name of the variable being analyzed). The default jobID is B=n^[exp],k=[c]x[-permute].

1.1 Example

```
java -jar MINE.jar "path/to/data.txt"  0  cv=0.1  exp=0.6  c=10  fewBoxes
```

This will run MINE on the file path/to/data.txt. The only variable pairs that will be analyzed are the first variable against the rest of the variables. Also, a variable pair will be ignored if less than 10% of the records have values for both the variables in question. The program will use $B(n) = n^{0.6}$ and will have the maximal number of clumps allowed be $k = 10x$ when attempting to draw a grix with x columns. Two output files will be created:

- 'path/to/data.txt,fewBoxes,Results.csv', which contains the results of the analysis, and
- 'path/to/data.txt,fewBoxes,Status.txt', which contains the name of the variable being analyzed while MINE runs.