Github Readme File to:
Searching for best lower dimensional visualization angles for high dimensional data

Wanli Zhang[1][¶] and Yanming Di[1][¶]

[1]Department of Statistics, Oregon State University,
Corvallis, Oregon, United States of America

[¶] These authors contributed equally to this work.

The folder **Code** contains the R code used by the authors to produce the results presented in the article.

(1) **data generation.R**: Produces the relative expression levels from raw read counts

(2) **figure_ code.R**: Generate all figures in the article

(3) **functions.R**: Contains all functions necessary for simulation and data analysis

(4) **PCA cross_ validation.R**: Generate results for PCA cross-validation

(5) **plot functions.R**: Contains functions necessary for generating graphs

(6) **run simulations.R**: Wrapper functions for generating simulation results

(7) **summarize simulation results.R**: Code for integrating simulation results into single files

(8) **Permutation_ code.R**: Permutation test with the purpose of calculating empirical p-values for feature subsets

(9) **sample_ size_ simulations.R**: Simulations designed to investigate variation and biasedness of GSI produced by RCV when sample size is large

(10) **FisherLDA.R**: Calculate GSI with LDA as classifier

(11) **tspair.R**: Use Top Scoring Pair to find top ranked feature pair

Folder **Data** contains the raw read counts as well as expression levels for genes involved in the study. See **data generation.R** for more details.

Folder **Graphs** contain all figures used in this article.

Folder **Output** contain the analysis output from the data generation process.

Folder **Results** contain repeated cross-validation results for 2-, 3-, and 4-dimensional feature subsets.