Kratos: Context-aware cell type classification and interpretation  
using joint dimensionality reduction and clustering

Author: Zihan Zhou, Zijia Du, Somali Chaterji

Requirement: Tensorflow 2.3.0

For explainer:

Default run – “python kratos.py”

Parameters:

--input\_dir : direction to the input dataset

--steps : the number of iterations of the neural network

--lamda : the coefficient on the regularization term

--lr : learning rate

--max\_iter : maximum iteration of the cw algorithm

--margin : the alpha term

--label : the label of the dataset

--use\_abs : whether to use absolute value

Example run on PBMC dataset:

python kratos.py --input\_dir ../../data/PBMC --steps 50 --lamda 100 --lr 0.001 --max\_iter 500 --margin 0 --label PBMC --use\_abs 1

The code will output gene relevance in npy format and gene importance score in csv format, e.g. ‘genes\_cw\_onevsrest\_lamda100.0\_iter500\_lr0.001\_clustid1\_abs0.csv’. The csv files will be used for next step. The clustering metrics and visualization results will be saved in the ‘validation’ folder.

For analyzer:

Default run – “python analyzer.py”

Parameters:

--input\_dir : direction to the dataset

--gene\_relavance\_url : direction to the gene score csv file

--target\_clustid : id of the target cluster

--use\_abs : whether to use absolute value when ordering the gene scores

Example run on PBMC dataset:

python analyzer.py --input\_dir ./data/PBMC --gene\_relevance\_url ./results/PBMC/results\_lr0.001\_steps50\_lamda100.0/genes\_cw\_onevsrest\_lamda100.0\_iter500\_lr0.001\_clustid1\_abs0.csv --target\_clustid 1 --use\_abs 1

The code will output the auroc and correlation using different subset of genes as csv format. Run through each cluster and rearrange the auroc and correlation files into this format and save as ‘analyze\_pred.csv’ and ‘analyze\_corr.csv’: Table

Description automatically generated

Table

Description automatically generated

Plot the final comparisons:

Default run – “python compare\_plot.py”

Parameters:

--input\_dir : direction to the dataset

Example run on PBMC dataset:

python compare\_plot.py --input\_dir ./results/PBMC/results\_lr0.001\_steps50\_lamda100.0

This code will output the comparison plots and save as pdf files.

Datasets:

PBMC: <https://noble.gs.washington.edu/proj/ACE/PBMC.zip>

Baron: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE84133>