RBCDE Documentation

Release 1.0.0

Krzysztof Polanski

Jan 17, 2020

Contents

rbcde.**RBC**(*adata*, *clus_key='leiden'*, *layer=None*, *use_raw=False*)

Compute the rank-biserial correlation coefficient for each gene in each cluster. The results can be subsequently turned into a marker list via the helper function `rbcde.filter_markers()`. The primary output is stored as part of either *.var* or *.raw.var*, depending on whether *.raw* data is used.

The rank-biserial correlation coefficient (Cureton, 1956) can be used as an effect size equivalent of the Wilcoxon test (Kerby, 2014), which in turn was deemed to perform well on single cell data problems (Soneson, 2018). Using effect size analyses is recommended for problems with large population sizes (Sullivan, 2012).

adata [`AnnData`] Needs per cell normalised data stored somewhere in the object (as either sparse or dense), and the desired clustering/grouping vector included in *.obs*.

clus_key [`str`, optional (default: "leiden")] The name of the *.obs* column containing the clustering/grouping.

layer [`str` or `None`, optional (default: `None`)] If specified, take the expression data from the matching `.layers` field. Overrides `use_raw` if provided.

use_raw [`bool`, optional (default: `False`)] If no `layer` was specified and this is set to `True`, take the data from the `.raw` field of the object. Store results in `.raw.var` to match dimensionality.

rbcde.**filter_markers**(*adata*, *thresh=0.5*, *use_raw=False*)

Filter the rank-biserial correlation coefficients computed with `rbcde.RBC()` to a list of markers for each cluster, provided as a data frame and a Scanpy plotting compatible `var_names` cluster marker dictionaty. Returns those two objects, in this order.

adata [`AnnData`] Needs to have been processed with `rbcde.RBC()`.

thresh [`float`, optional (default: 0.5)] The threshold value used to call markers. Literature critical values can be used.

use_raw [`bool`, optional (default: `False`)] Set this to `True` if the raw data was used for the computation so that the results can be retrieved from the correct field of the object.

CHAPTER 1

rbcde.matrix

rbcde.matrix.**RBC**(*data*, *clusters*, *vars*)

Compute the rank-biserial correlation coefficient for each gene in each cluster. The results can be subsequently turned into a marker list via the helper function `rbcde.matrix.filter_markers()`. Returns a data frame with the coefficient value for each gene in each cluster.

The rank-biserial correlation coefficient (Cureton, 1956) can be used as an effect size equivalent of the Wilcoxon test (Kerby, 2014). Using effect size analyses is recommended for problems with large population sizes (Sullivan, 2012).

data [`np.array` or `scipy.sparse`] Per cell normalised, if using single cell count data. Variables as rows, observations as columns.

clusters [`np.array` or `list`] A vector of cluster/group assignments for each observation.

vars: `np.array` or `list` A vector of variable names, for output generation purposes.

rbcde.matrix.**filter_markers**(*results*, *thresh=0.5*)

Filter the rank-biserial correlation coefficients computed with `rbcde.matrix.RBC()` to a list of markers for each cluster. Returns a data frame of the computed markers.

results [`pd.DataFrame`] The output of `rbcde.matrix.RBC()`.

thresh [`float`, optional (default: 0.5)] The threshold value used to call markers. Literature critical values can be used.

Python Module Index

r

Index

F

R