pyALRA

Alexandre LANAU

Jun 23, 2025

CONTENTS:

PYALRA.CORE MODULE

1.1 pyALRA.core.alra

pyALRA.core.**alra**(*A_norm*, *k=0*, *q=12*, *quantile_prob=0.001*, *random_state=1*, *svd_type=None*)

   Adaptive thresholded low-rank approximation (ALRA) for imputation of sparse data.

   Parameters

   - **A_norm** (`ndarray or sparse matrix`) – Normalized input matrix.
   - **k** (`int, optional`) – Rank for approximation; if 0, automatically chosen (default 0).
   - **q** (`int, optional`) – Number of power iterations for randomized SVD (default 12).
   - **quantile_prob** (`float, optional`) – Quantile threshold for adaptive thresholding (default 0.001).
   - **random_state** (`int, optional`) – Random seed for reproducibility (default 1).
   - **svd_type** (`str or None, optional`) – SVD method: 'truncated' or None for randomized SVD.

   Returns

   Dictionary containing: - 'A_norm_rank_k': low-rank approximation matrix (rank k) - 'A_norm_rank_k_cor': thresholded low-rank matrix - 'A_norm_rank_k_cor_sc': scaled and thresholded matrix (final imputed matrix)

   Return type

   dict

1.2 pyALRA.core.choose_k

pyALRA.core.**choose_k**(*A_norm*, *K=100*, *thresh=6*, *noise_start=80*, *q=12*, *random_state=1*, *svd_type=None*)

   Select the rank k for low-rank approximation based on singular value gap statistics.

   Parameters

   - **A_norm** (`ndarray or sparse matrix`) – Normalized input matrix.
   - **K** (`int, optional`) – Maximum number of singular values to consider (default 100).
   - **thresh** (`float, optional`) – Threshold on number of standard deviations to detect significant singular value gap (default 6).
   - **noise_start** (`int, optional`) – Index to start noise singular values (default 80).
   - **q** (`int, optional`) – Number of power iterations for randomized SVD (default 12).

- **random_state** (`int, optional`) – Random seed for reproducibility (default 1).

- **svd_type** (`str or None, optional`) – SVD method: 'truncated' or None for randomized SVD.

Returns

Dictionary with keys: - 'k': selected rank - 'num_of_sds': array of standardized singular value differences - 'd': singular values array

Return type

dict

1.3 pyALRA.core.randomized_svd_py

pyALRA.core.**randomized_svd_py**(*A*, *K*, *q*, *random_state*, *svd_type=None*)

Perform SVD with an option for randomized or truncated SVD.

Parameters

- **A** (`ndarray or sparse matrix`) – Input data matrix to decompose.

- **K** (`int`) – Number of singular values and vectors to compute.

- **q** (`int`) – Number of power iterations (only applicable for randomized SVD).

- **random_state** (`int`) – Random seed for reproducibility.

- **svd_type** (`str or None, optional`) – If 'truncated', use TruncatedSVD; otherwise, use randomized SVD.

Returns

- **U** (*ndarray*) – Left singular vectors.

- **Sigma** (*ndarray*) – Singular values.

- **VT** (*ndarray*) – Right singular vectors transposed.

1.4 pyALRA.core.normalize_data

pyALRA.core.**normalize_data**(*A*)

Normalize data by library size and log-transform.

Parameters

A (`ndarray or sparse matrix`) – Input matrix with cells as rows and genes as columns.

Returns

Normalized and log-transformed matrix.

Return type

ndarray

INDEX