Documentation for Python Scripts

This document provides detailed documentation for the Python scripts: mds.py, pca1.py, pca2.py, and pca3.py. Each script is designed to perform specific data processing tasks, primarily focusing on dimensionality reduction techniques. The document outlines the purpose, functions, and usage of each script.

1. pca1.py

Purpose

The pca1.py script performs Principal Component Analysis (PCA) without mean subtraction. It reduces the dimensionality of the data by projecting it onto the principal components.

Functions

check_input(input_file)

This function checks the input CSV file for validity, similar to the mds.py script.

pca_without_mean_subtraction(data, n_components=2)

This function computes the covariance matrix of the data and performs eigendecomposition to obtain the principal components.

main(input_file, output_file)

The main function reads the input data, performs PCA, and saves the reduced data to the output file.

Usage

To execute the script, use the following command:
python pca1.py <input_file> <output_file>

2. pca2.py

Purpose

The pca2.py script performs PCA with mean subtraction. It centers the data before computing the principal components, which often leads to better results.

Functions

check_input(input_file)

This function checks the input CSV file for validity, similar to the previous scripts.

pca_with_mean_subtraction(data, n_components=2)

This function centers the data by subtracting the mean, computes the covariance matrix, and performs eigendecomposition.

main(input_file, output_file)

The main function reads the input data, performs PCA with mean subtraction, and saves the reduced data to the output file.

Usage

To execute the script, use the following command:
python pca2.py <input_file> <output_file>

3. pca3.py

Purpose

The pca3.py script performs a normalized PCA with mean subtraction. It normalizes the data before computing the principal components, which can be useful for certain datasets.

Functions

check_input(input_file)

This function checks the input CSV file for validity, similar to the previous scripts.

normalized_pca_with_mean_subtraction(data, n_components=2)

This function normalizes the data, centers it by subtracting the mean, and computes the principal components.

main(input_file, output_file)

The main function reads the input data, performs normalized PCA with mean subtraction, and saves the reduced data to the output file.

Usage

To execute the script, use the following command:
python pca3.py <input_file> <output_file>

4. mds.py

Purpose

The mds.py script implements Multidimensional Scaling (MDS) using an exponential scaling approach. It reduces the dimensionality of input data while preserving the pairwise distances between data points.

Functions

check_input(input_file)

This function reads the input CSV file and checks for the validity of the data. It ensures the data has at least two columns and rows, and contains only numeric values.

mds_exponential(data, n_components=2, alpha=1.0)

This function performs the MDS transformation on the input data. It computes pairwise squared distances, applies custom scaling, and performs eigendecomposition to obtain reduced coordinates.

main(input_file, output_file, alpha)

The main function orchestrates the reading of input data, performs MDS, and saves the reduced data to the specified output file.

Usage

To execute the script, use the following command:

python mds.py <input_file> <output_file> <alpha>