The k-means clustering algorithm is a commonly used method for grouping *n* individual data points into *k* clusters. is a multi-variate statistical analysis that reduces the high-dimensional matrix of correlated, time-varying signals into a low-dimensional and statistically uncorrelated set of principal components (PCs). These PCs explain the variance found in the original signals and represent the most important features of the data, e.g., the overall magnitude or the shape of the time series at a particular point in the stride cycle. The value of each particular subject's score for the individual PCs represents how strongly that feature was present in the data.

The utility of clustering

When analysing biomechanical signals, we often realize that a number of individual traces are similar. It can be useful to describe these traces as belonging to the same group, or cluster. This potentially allows us to simplify our analysis or to pick a single trace as being "representative" of the whole cluster. Because clustering is an unsupervised learning technique, it does not require any specific knowledge or set of training labels from the user. This, in turn, makes clustering useful for data exploration.

Performing k-means clustering

Inspect3D allows users to apply the k-means clustering algorithm to the results of PCA. The dimensionality of the data space is the number of principal components and the user specifies the number of clusters to be found - this is the parameter *k*.

- In the **PCA** menu there is the option to perform quality assurance (QA) on the results;
- Selecting **Run QA with PCA** brings up a window with multiple options for QA;
- The **K-Means** tab allows the user to specify parameter values for the algorithm and then run it on PCA results.

Reference

The k-means clustering algorithm is more than 50 years old and is described in almost every textbook on data analysis and machine learning. Inspect3D specifically implements the k-means++ algorithm, which optimizes how the initial cluster centres are chosen.

Arthur D. and Vassilvitskii S. (2007). k-means++: the advantages of careful seeding. Proceedings of the eighteenth annual ACM-SIAM symposium on discrete algorithms. Society for Industrial and Applied Mathematics Philadelphia, PA, USA. pp. 1027–1035. **Abstract** The k-means method is a widely used clustering technique that seeks to minimize the average squared distance between points in the same cluster. Although it offers no accuracy guarantees, its simplicity and speed are very appealing in practice. By augmenting k-means with a simple, randomized seeding technique, we obtain an algorithm that is O(log k)-competitive with the optimal clustering. Experiments show our augmentation improves both the speed and the accuracy of k-means, often quite dramatically.

From:
https://has-motion.com/wiki/ - **Wiki Documentation Site**

Permanent link:
https://has-motion.com/wiki/doku.php?id=inspect3d:documentation:knowledge_discovery:k-means_clustering&rev=1719432816

Last update: **2024/06/26 20:13**