kmeans

NAME

kmeans - A clustering algorithm commonly used in EDA (exploratory data analysis).

SYNOPSIS

```
#include <frovedis/ml/clustering/kmeans.hpp>
```

```
rowmajor_matrix_local<T>
frovedis::kmeans (crs_matrix<T,I,O>& samples,
      int k,
      int iter,
      T eps,
      long seed = 0)
```

```
std::vector<int>
frovedis::kmeans_assign_cluster (crs_matrix_local<T,I,O>& mat,
          rowmajor_matrix_local<T>& centroid)
```

DESCRIPTION

Clustering is an unsupervised learning problem whereby we aim to group subsets of entities with one another based on some notion of similarity. K-means is one of the most commonly used clustering algorithms that clusters the data points into a predefined number of clusters (K).

Detailed Description

frovedis::kmeans()

Parameters
samples: A `crs_matrix<T,I,O>` containing the sparse data points
k: An integer parameter containing the number of clusters
iter: An integer parameter containing the maximum number of iteration count
eps: A parameter of T type containing the epsilon value
seed: A parameter of long type containing the seed value to generate the random rows from the given data samples (Default: 0)

Purpose
It clusters the given data points into a predefined number (k) of clusters.
After the successful clustering, it returns the k centroids of the cluster.

Return Value
After the successful ustering it returns the centroids of the type `rowmajor_matrix_local<T>`, where each column shows each centroid vector.

frovedis::kmeans_assign_cluster()

Parameters
mat: A `crs_matrix_local<T,I,O>` containing the new sparse data points to be assigned to the cluster
centroid: A `rowmajor_matrix_local<T>` contaning the centroids

Purpose
After getting the centroids from kmeans(), they can be used to assign data to the closest centroid using kmeans_assign_cluster().

Return Value
It returns a `std::vector<int>` containing the assigned values.