File Organization

2021168848/
        KMeans
        EMAlgorithm

Section 1: KMeans

Contains the Class Files, Source Files, Data Files.

The files are organized as follows

-bin/

-        KMeans.class

- src/

        -KMeans.java

- outputImages/

- .settings
- .classpath

- result.pdf

Result.pdf contains the results.

Outputimages – contains the output images.
Running the program

The program can be made to run as follows.
Navigate to the bin folder in the command prompt and use the following command

`>java KMeans InputImageName clusterSize OutPutImageName`

Example command

`>java Koala.jpg 2 koalaop.jpg`

The program expects the three arguments.

1. InputImageName : Name of the input image

2. clusterSize : size of the cluster. Larger the number better is the quality. Start with 2 and a cluster of size 20 gives a better quality image.

3. OutPutImageName: What should be the output name of the image? That name goes here.

All the source files are found inside the src folder.

It is recommended to follow the same directory structure to avoid any confusions. If there is a case to add a new data set add it under "bin" folder and follow the same directory structure as other datasets have.

File Descriptions

KMeans.java

      Contains various methods that performs KMeans clustering on the input data.

`kmeans( int[] pixels, int k ) –  Gets the pixels and cluster size and performs kMeans on the given input data.`

`isConverged( int[] previousCentroids, int[] currentCentroids )  - Checks if KMeans is converged.`

`calculatePixelDistance( int pixelA, int pixelB ) – Calculates the distance between pixel A and pixel B`

Section 2: EMAlgorithm

-bin/

- DataSetReader.class
- EM.class
- Main.class
- em_data.txt

- src/

- DataSetReader.java
- EM.java
- Main.java
- em_data.txt

- .settings

- .classpath

- result.pdf

All the source files are found inside the src folder

It is recommended to follow the same directory structure to avoid any confusions. If there is a case to add a new data set add it under "bin" folder and follow the same directory structure as other datasets have.

Result.pdf contains the results.

The program can be made to run as follows:

Navigate to the bin folder in the command prompt and use the following command

>java Main inputFileName

Example command

>java em_data.txt

File Descriptions

DataSetReader.java

readDataFromTheFile(String fileName) – Read the data from the input and creates a array list of type double.

EM.java

loadDataFromTheFile(String fileName) – Invokes the readDataFromTheFile() method for creating the 1 Dimensional data from the file.

setNumOfClusters(int cluster) – Sets the number of the cluster.

init() – Starts the EM Algorithm which composes of two steps the E Step and M Step.

converged(double logliklyhood, double previousLogLiklyHood) – Checks if the algorithm has converged. Returns true if it has converged else returns false.

MStep() – Performs the Maximization of EM Algorithm. Update means for clusters, covariance matrix entries, and total probabilities for each class.

EStep() – Performs the Expected Step of the algorithm. Update probabilities of belonging to each class for each data point

calculateLogLikelyHood() – Returns the log likely hood for the given data

gaussian(double xi, double mean, double var) - Just a univariate Gaussian because the data is 1D

printData(String type) – Prints the value for Mean Matrix, Initial Variance Matrix and Probability Totals Per Cluster.

Main.java

Contains the main method which invokes other methods.