# Assignment 4

## K - means

1. K-means clustering is a common example of an exclusive clustering method where data points are assigned into K groups, where K represents the number of clusters based on the distance from each group's centroid.

2. The data points closest to a given centroid will be clustered under the same category.

3. A larger K value will be indicative of smaller groupings with more granularity, whereas a smaller K value will have larger groupings and less granularity.

4. K-means clustering is commonly used in market segmentation, document clustering, image segmentation, and image compression.

**How to Implement K-Means Clustering?**

1. Choose the number of clusters k.
2. Select k random points from the data as centroids.
3. Assign all the points to the closest cluster centroid.
4. Recompute the centroids of newly formed clusters.
5. Repeat steps 3 and 4 until convergence or fix number of steps.

Implement the above algorithm in C++.

## LGB Algorithm

The Linde-Buzo-Gray (LBG) algorithm is a vector quantization algorithm that is commonly used for signal compression and image processing. Here are the steps for the LBG algorithm:

Initialize the codebook: Start with a small codebook of size M, where each codeword is randomly generated or selected from the training data.

Split each codeword into two: Divide each codeword into two equal parts and replace it with two new codewords that are identical to the original codeword except for one part, which is either increased or decreased by a small amount. This doubles the size of the codebook.

Iterate until convergence: Repeat the following steps until the codebook stops changing:

a. Assign each training vector to its closest codeword in the current codebook.

b. Compute the centroid of each cluster: For each codeword in the current codebook, compute the mean of the training vectors that belong to the cluster associated with that codeword.

c. Update the codewords: Replace each codeword in the current codebook with its two nearest centroids, which become the new codewords for the next iteration. If there are ties, choose the centroid with the lowest index.

Output the final codebook: After convergence, output the final codebook, which consists of the M closest centroids to each of the original M codewords.

Encoding and decoding: To encode a new signal, find the closest codeword in the codebook and replace it with the index of that codeword. To decode the signal, replace each index with the corresponding codeword from the codebook.

The LBG algorithm can be sensitive to the choice of initial codebook, so it is often run multiple times with different random initializations and the best codebook is chosen based on some criterion, such as the distortion or the bit rate of the compressed signal.

Implement the above algorithm in C++. .

**Instructions to students**

1. Please make sure that there is no plagiarism in the implemented code.
2. Please submit your assignment before 10 A.M. on Wednesday.