Details of the k-means Algorithm

k-means algorithm

Input: dataset $\mathcal{D} = \{\mathbf{x}_i\}$ and number k of clusters
Output: codebook $\mathcal{C} = \{\mathbf{z}_j\}$ and assignment function y

while termination criterion is not met **do**
  // encoding/assignment step
  for each instance $\mathbf{x}_i$ **do**
    $y(\mathbf{x}_i) = \arg\min_{j=1\ldots k} d(\mathbf{x}_i, \mathbf{z}_j)$
  end for

  // decoding/codebook update step
  for each centroid $\mathbf{z}_j$ **do**
    $\mathbf{z}_j = \frac{1}{|\mathcal{C}_j|} \sum_{i \in \mathcal{C}_j} \mathbf{x}_i$
  end for
end while

termination: # of iterations, change of successive codebooks / $J(\mathcal{C})$ values