LBG Algorithm

LBG algorithm is like a K-means clustering algorithm which takes a set of input vectors $S = \{\mathbf{x}_i \in R^d |\ i = 1, 2,\ \ldots, n\}$ as input and generates a representative subset of vectors $C = \{\mathbf{c}_j \in R^d |\ j = 1, 2,\ \ldots, K\}$ with a user specified $K << n$ as output according to the similarity measure. For the application of Vector Quantization (VQ), $d = 16$, $K = 256$ *or* 512 are commonly used.

LBG Algorithm

1. Input training vectors $S = \{\mathbf{x}_i \in R^d |\ i = 1, 2,\ \cdots, n\}$.
2. Initiate a codebook $C = \{\mathbf{c}_j \in R^d |\ j = 1, 2,\ \cdots, K\}$.
3. Set $D_0 = 0$ and let $k = 0$.
4. Classify the n training vectors into K clusters according to $\mathbf{x}_i \in S_q$ if $\|\mathbf{x}_i - \mathbf{c}_q\|_p \ \leq\ \|\mathbf{x}_i - \mathbf{c}_j\|_p$ for $j \neq q$.
5. Update cluster centers $\mathbf{c}_j,\ j = 1, 2,\ \cdots, K$ by $\mathbf{c}_j = \frac{1}{|S_j|} \sum_{\mathbf{x}_i \in S_j} \mathbf{x}_i$.
6. Set $k \ \leftarrow\ k + 1$ and compute the distortion $D_k = \sum_{j=1}^{K} \sum_{\mathbf{x}_i \in S_j} \|\mathbf{x}_i - \mathbf{c}_j\|_p$.
7. If $(D_{k-1} - D_k)/D_k \ >\ \epsilon$ (*a small number*), repeat steps $4 \sim 6$.
8. Output the codebook $C = \{\mathbf{c}_j \in R^d |\ j = 1, 2,\ \cdots, K\}$,

The convergence of LBG algorithm depends on the initial codebook C, the distortion D_k, and the threshold ϵ, in implementation, we need to provide a maximum number of iterations to guarantee the convergence.