k-means Algorithm: Decoding Step

encoder/assignment is known, what is the best decoder/codebook ?

Optimal solution for the decoding step

$$\mathbf{x}_i \longrightarrow \boxed{\text{encoder}} \longrightarrow \text{index } y(\mathbf{x}_i) \longrightarrow \boxed{\textbf{decoder}} \longrightarrow \text{centroid } \mathbf{z}_{y(\mathbf{x}_i)}$$

second step: decoder which minimise the training reconstruction error

$$J(\mathcal{C}) = \frac{1}{n} \sum_{i=1}^{n} d(\mathbf{x}_i, \mathbf{z}_{y(\mathbf{x}_i)})^2 = \sum_{j=1}^{k} \left(\frac{|\mathcal{C}_j|}{n} \right) \left(\frac{1}{|\mathcal{C}_j|} \sum_{i \in \mathcal{C}_j} d(\mathbf{x}_i, \mathbf{z}_{y(\mathbf{x}_i)})^2 \right)$$

solution: move centroid $\mathbf{z}_j$ to the center of gravity of cluster $\mathcal{C}_j$

$$\mathbf{z}_j = \frac{1}{|\mathcal{C}_j|} \sum_{i \in \mathcal{C}_j} \mathbf{x}_i$$