k-means Algorithm: Encoding Step

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

Optimal solution for the encoding step

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

first step: encoder which minimise the training reconstruction error

$$J(\mathcal{C}) = \frac{1}{n} \sum_{i=1}^{n} d(\mathbf{x}_i, \mathbf{z}_{y(\mathbf{x}_i)})^2$$

solution: assign $\mathbf{x}_i$ to the closest centroid $\mathbf{z}_{y(\mathbf{x}_i)}$

$$y(\mathbf{x}_i) = \arg \min_{j=1 \ldots k} d(\mathbf{x}_i, \mathbf{z}_j)$$