Derivation of the k-means Algorithm

Encoding-decoding view

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

goal : encoder and decoder which minimise the training reconstruction error

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

Approximate solution with the k-means algorithm

- no analytical solution to the encoder-decoder problem
- k-means algorithm: iterative, greedy algorithm
- start from an initial decoder (codebook), then improve it