Derivation of the k-means Algorithm

Notations

- $\mathcal{C} = \{\mathbf{z}_j\}$: codebook of centroids
- $y(\mathbf{x})$ = index of the centroid to which is assigned instance $\mathbf{x}$

Objective function

minimise reconstruction error with the codebook of centroids

$$J(\mathcal{C}) = \int_{\mathbf{x}} p(\mathbf{x}) d(\mathbf{x}, \mathbf{z}_{y(\mathbf{x})})^2 d\mathbf{x}$$

which is approximated by the training reconstruction error

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