

Problem. Optimize Loss function of K-Means algorithm

Solution.

We will start with an unsupervised learning (clustering) problem. Given a dataset $\{x_1, \dots, x_N\}$, each $x_i \in R^D$, partition the dataset into K clusters.

We denote cluster centers: $\mu_k \in R^D$. Use binary r_{nk} , 1 if point n is in cluster k , 0 otherwise.

Find $\{\mu_k\}$, $\{r_{nk}\}$ to minimize distortion measure:

$$J = \sum_{n=1}^N \sum_{k=1}^K r_{nk} \|x_n - \mu_k\|^2 \quad (1)$$

First, minimize distortion measure J wrt cluster membership variables (label vector) r_{nk} with the μ_k held fixed. Terms for different data points x_n are independent, for each data point set r_{nk} to minimize

$$\sum_{k=1}^K r_{nk} \|x_n - \mu_k\|^2 \quad (2)$$

Simply set $r_{nk} = 1$ for the cluster center μ_k with smallest distance. So x_n belongs to the cluster that has the smallest distance to x_n .

Second, minimize J wrt the cluster centers μ_k (minimize wrt each μ_k separately). Take derivative, set to zero:

$$2 \sum_{n=1}^N r_{nk} (x_n - \mu_k) = 0 \quad (3)$$

$$\Leftrightarrow \mu_k = \frac{\sum_n r_{nk} x_n}{\sum_n r_{nk}} \quad (4)$$

μ_k equal to the mean of all data points x_n assigned to cluster k .