K Means

Clustering

- Given: N unlabeled examples $\{x_1,\cdot\cdot\cdot,x_N\}$ no. of desired partitions K.
- ullet Goal: : Group the examples into K "homogeneous" partitions.

Given a set of observations $X=\{x_1,x_2,\cdot\cdot\cdot,x_N\}(x_i\in\mathbb{R}^D)$, partition the N observations into K sets ($K\leq N$) $\{\mathcal{C}_k\}_{k=1,\cdot\cdot\cdot,K}$ such that the sets minimize the within-cluster sum of squares:

$$\arg\min_{\mathcal{C}_k} \sum_{i=1}^k \sum_{x \in \mathcal{C}_i} ||x - \mu_i||^2$$

where μ is the mean of points in set \mathcal{C}_i .

K Means Algorithm

Algorithm

• (Re)-Assign each example x_i to its closest cluster center (based on the smallest Euclidean distance)

$$\mathcal{C}_{k} = \left\{ \boldsymbol{x}_{i} \mid \left\| \boldsymbol{x}_{i} - \boldsymbol{\mu}_{k} \right\|^{2} \leq \left\| \boldsymbol{x}_{i} - \boldsymbol{\mu}_{k'} \right\|^{2}, \text{ for } \forall k' \neq k \right\}$$

- (\mathcal{C}_k is the set of examples assigned to cluster k with center μ_k) - Update the cluster means

$$\mu_k = \operatorname{mean}\left(\mathcal{C}_k\right) = \frac{1}{|\mathcal{C}_k|} \sum_{x \in \mathcal{C}_k} x$$

• Let $z_{i,k}$ be an indicator

$$z_{i,k} = \left\{ \begin{array}{ll} 1, & x_i \in \mathcal{C}_k \\ 0, & otherwise \end{array} \right.$$

- and $z_i = [z_{i,1}, \cdots, z_{i,k}]^T$ represents the one-hot encoding of $x_i.$
- $\begin{array}{l} \bullet \text{ The loss is } L(\mu, \mathbf{X}, \mathbf{Z}) = \sum_{i=1}^{N} \sum_{k=1}^{K} z_{i,k} \left\| x_i \mu_k \right\|^2 = \|\mathbf{X} \mathbf{Z}\mu\|^2 \\ \bullet \text{ where } \mathbf{X} \in \mathbb{R}^{N \times D}, \mathbf{Z} \in \mathbb{R}^{N \times K} \text{ , } \mu \in \mathbb{R}^{K \times D} \end{array}$

Limitations

- · Makes hard assignments of points to clusters
- · Works well only is the clusters are roughly of equal sizes
- · K-means also works well only when the clusters are round-shaped and does badly if the clusters have non-convex shapes

Kernel K Means

• Basic idea: Replace the Euclidean distance/similarity computations in K-means by the kernelized versions

$$\begin{split} d\left(x_{i}, \mu_{k}\right) &= \left\|\phi\left(x_{i}\right) - \phi\left(\mu_{k}\right)\right\| \\ \left\|\phi\left(x_{i}\right) - \phi\left(\mu_{k}\right)\right\|^{2} &= \left\|\phi\left(x_{i}\right)\right\|^{2} + \left\|\phi\left(\mu_{k}\right)\right\|^{2} - 2\phi\left(x_{i}\right)^{T}\phi\left(\mu_{k}\right) \\ &= k\left(x_{i}, x_{i}\right) + k\left(\mu_{k}, \mu_{k}\right) - 2k\left(x_{i}, \mu_{k}\right) \end{split}$$