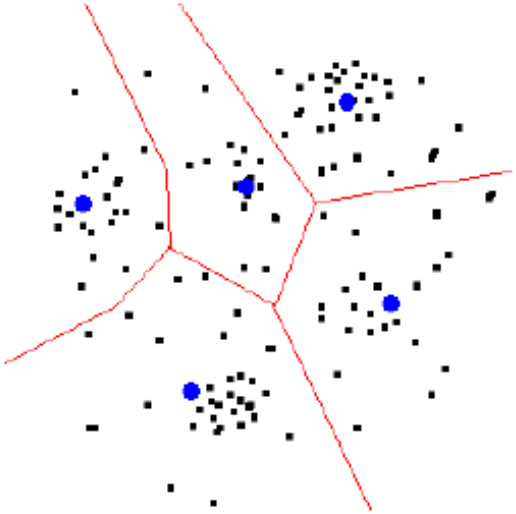
K-means

Clustering

- Unsupervised Learning
- Divide data into clusters

- What is K-means
  - Find a set of K centers given set of data points



Algorithm

$$\mathcal{J} = \sum_{j=1}^{K} \sum_{t \in \mathcal{C}_j} \left\| \boldsymbol{x}_t - \boldsymbol{\mu}_j \right\|^2$$

① Initialize K centers with K randomly selected data

② Compute the distance between each K centers and each data point (N * K matrix)

③ Determine the closest center for each data point

④ Compute new cluster center for all clusters

Normalized Mutual Information

Normalized mutual information (NMI) between $\mathcal{C}$ and $\widetilde{\mathcal{C}}$ which correspond to the set of estimated clusters and the set of ground truth clusters, respectively.

$$\boxed{\mathrm{NMI}(\mathcal{C}, \widetilde{\mathcal{C}}) = \frac{I(\mathcal{C}, \widetilde{\mathcal{C}})}{\left[H(\mathcal{C}) + H(\widetilde{\mathcal{C}})\right]/2},}$$

where the mutual information is calculated as

$$\begin{aligned} I(\mathcal{C}, \widetilde{\mathcal{C}}) &= \sum_i \sum_j p(C_i, \widetilde{C}_j) \log_2 \frac{p(C_i, \widetilde{C}_j)}{p(C_i)p(\widetilde{C}_j)} \\ &= \sum_i \sum_j \frac{|C_i \cap \widetilde{C}_j|}{N} \log_2 \frac{N|C_i \cap \widetilde{C}_j|}{|C_i||\widetilde{C}_j|}, \end{aligned}$$

and the entropy is computed as

$$H(\mathcal{C}) = -\sum_j p(C_j) \log_2 p(C_j) = -\sum_j \frac{C_j}{N} \log_2 \frac{C_j}{N}.$$

Alternative Algorithm

- Introduce responsibilities which are indicator variables,

$$r_{jt} = \begin{cases} 1, & \text{if} \quad \arg\min_k \|\boldsymbol{x}_t - \boldsymbol{\mu}_k\|^2 = j, \\ 0 & \text{if} \quad \arg\min_k \|\boldsymbol{x}_t - \boldsymbol{\mu}_k\|^2 \neq j. \end{cases}$$

$$\mathcal{J} = \sum_{j=1}^{K} \sum_{t=1}^{N} r_{jt} \|\boldsymbol{x}_t - \boldsymbol{\mu}_j\|^2$$

- Algorithm
  - Assignment step: compute responsibilities
  - Update step: update centers

$$\boldsymbol{\mu}_j = \frac{\sum_{t=1}^{N} r_{jt} \boldsymbol{x}_t}{\sum_{t=1}^{N} r_{jt}}$$

Soft K-means Algorithm

$$\mathcal{J}_{SK} = \underbrace{\sum_{t=1}^{N}\sum_{j=1}^{K} r_{jt}\|\mathbf{x}_t - \boldsymbol{\mu}_j\|^2}_{\text{expected energy}} - \frac{1}{\beta}\underbrace{\sum_{t=1}^{N}\sum_{j=1}^{K} r_{jt}\log\frac{1}{r_{jt}}}_{\text{entropy}}$$

- Assignment step
  - Each data point $\mathbf{x}_t$ is given a soft degree of assignment to each of the centers. We call the degree to which $\mathbf{x}_t$ is assigned to cluster j, the responsibility

$$r_{jt} = \frac{\exp\left\{-\beta\left\|\mathbf{x}_t - \boldsymbol{\mu}_j\right\|^2\right\}}{\sum_l \exp\left\{-\beta\left\|\mathbf{x}_t - \boldsymbol{\mu}_l\right\|^2\right\}} \qquad \text{where} \quad \sum_j r_{jt} = 1$$

- Update step
  - The model parameters, the centers, are adjusted to match the sample means of the data points that they are responsible for

$$\boldsymbol{\mu}_j = \frac{\sum_{t=1}^{N} r_{jt}\mathbf{x}_t}{\sum_{t=1}^{N} r_{jt}}$$