Notes of Machine Learning
Unsupervised Learning

Xiangli Chen

Computer Science Department
University of Illinois at Chicago

July 16, 2017

Outline

Problem Formulation

K-means clustering problem: partition n observations into K sets ($K \leq n$) where $S = \{S_1, \ldots, S_k\}$ such that

$$\arg\min_S \sum_{k=1}^{K} \sum_{c_i \in S_k} ||x_i - \mu_k||_2^2$$

where μ_k is the mean of points in set S_k.
Remark

- NP hard problem
- Heurictic algorithm: K-means algorithm

K-means Algorithm

- Initialize K centers $\hat{\mu}^0 = (\hat{\mu}^0_1, \ldots, \mu^0_K)$
- Do
  - Classify (fix μ, optimize C): at iteration t, assign each point $i \in \{1, \ldots, n\}$ nearest center

$$\forall i \quad \hat{C}^t(i) = \arg\min_k ||x_i - \hat{\mu}^t_k||^2_2$$

  - Recenter (fix C, optimize μ)

$$\forall k \quad \hat{\mu}^{t+1}_k = \arg\min_\mu \sum_{i \in \{\hat{C}^t(i) = k\}} ||x_i - \mu||^2_2$$

- Termination: none of n objects change membership

Complexity

Note that

$$\forall k \quad \frac{\partial \sum_{i \in \{\hat{C}^t(i)=k\}} ||x_i - \mu||_2^2}{\partial \mu} = 0 \Rightarrow$$

$$\hat{\mu}_k^{t+1} = \frac{\sum_{i \in \{\hat{C}^t(i)=k\}} x_i}{|\hat{C}^t(i) = k|} = \bar{x}_{i\, i \in \{\hat{C}^t(i)=k\}}$$

Time complexity

- clarification $O(Kn)$
- recentering (compute the new mean) $O(n)$

Assume I iteractions, the time complexity is $O(IKn)$.

Number of Clusters

The dissimlarity

$$W_k = \sum_{k=1}^{K} \sum_{i \in \{\hat{C}^t(i)=k\}} ||x_i - \hat{\mu}_k||_2^2$$

In general, $\{W_1, W_2, \ldots, W_{K_{\max}}\}$ decreases with increasing K.
Intuition: spliting a natural group reduces the criterion $W_k - W_{k+1}$ less than partitioning the union of two well-separated groups into their proper constituents.
Assume there are actually K^* distinct groupings, there will be a sharp decrease of $W_k - W_{k+1}$ at $k = k^*$, i.e.,

$$\{W_k - W_{k+1} | k < k^*\} >> \{W_k - W_{k+1} | k \geq k^*\}$$

Gap Statistics

To be continued

Seed Choice

Multiple random selections

Forward stepsize assignment (choose datapoints to be centers)

- initialize x_{i_1}
- do

  choose new center x_{i_m}

  for $i_m \in \{1, \dots, n\} - \{i_1, \dots, i_{m-1}\}$

  $\forall i \quad C(i) = \arg\min_m ||x_i - x_{i_m}||_2^2$

  $W_{i_m} = \sum_{k=1}^{m} \sum_{i \in \{C(i)=m\}} ||x_i - x_{i_k}||_2^2$

  $i_m = \arg\min_{i_m} W_{i_m}$
- $m = K$

Relation to Mixture Model and EM

To be continued

Hierarchical Clustering

To be continued