**Machine Learning**, Spring 2019
EN 2105
MWF 9-9:50am

Chao Lan
EN 4084A
clan@uwyo.edu

# K-means Clustering

Clustering is the task of grouping instances so that instances assigned to the same group are similar (and those assigned to different groups are dissimilar). The similarity between instances in the same group is intra-cluster similarity, and that between instances in differen groups is inter-cluster similarity. There are many ways to measure similarity (e.g., L2 or L1 distance).

The K-means clustering method finds a grouping assignment of high intra-cluster similarity. It measures similarity using L2 distance. The number of desired clusters $K$ is a hyper-parameter. Let $C_k$ be the set of instances assigned to cluster $k$. Let $\mu_k$ be the centroid of $C_k$. K-means finds assignments $\{(C_k, \mu_k)\}_k$ that minimizes the following objective function

$$J(C_k, \mu_k) = \sum_{k=1}^{K} \sum_{x \in C_k} ||x - \mu_k||^2. \tag{1}$$

Solving (1) directly is computationally intractable. One often applies the Lloyd's Algorithm to alternately optimize $C_k$ and $\mu_k$ until convergence.

---

**Algorithm 1** The Lloyd's Algorithm for K-means Clustering

---

    **Input:** a sample $S$, the number of clusters $K$
    **Initialization:** (randomly) initialize $K$ centroids $\mu_1, \ldots, \mu_K$
    **while** $C_k$ is still changing **do**
        1: for each $x$, assign it to cluster $k$ if its nearest centroid is $\mu_k$, i.e.,

$$k = \arg\min_{j=1,\ldots,K} ||x - \mu_j||. \tag{2}$$

        2: for each $\mu_k$, update it based on the new assignment $C_k$ from (2), i.e.,

$$\mu_k = \frac{1}{|C_k|} \sum_{x \in C_k} x. \tag{3}$$

    **end while**
    **Output:** converged cluster assignments $C_1, \ldots, C_K$

---

[*Discussion*] Give a demo of the Lloyd's algorithm.

K-means is guaranteed to converge, because $J$ decreases as (2) and (3) alternates. In fact, one can verify that (i) fixing $\mu_k$'s, (2) minimizes $J$ and (ii) fixing $C_k$'s, (3) minimizes $J$. But K-means may converge to local minimum and is sensitive to centroid initialization. In practice, one can run K-means multiple times and pick up the assignment with the smallest $J$.

[*Exercise*] Verify facts (i) and (ii).