CS150 Discussion 12

Jiachun Jin

K-means Clustering

The optimization problem

notations

- Data points: $\mathbf{x}_i \in \mathbb{R}^D, i = 1, 2, \ldots, N$
- K: number of clusters
- Centers of clusters: $\boldsymbol{\mu}_k, k = 1, 2, \ldots, K$
- indicator variables: $r_{nk} = \mathbb{I}[\mathrm{x}_i \text{ is assigned to the } k^{th} \text{cluster}] \in \{0, 1\}$

goal

Find an assignment of data points to clusters, as well as a set of vectors $\{\boldsymbol{\mu}_k\}$, such that the sum fo the squares of the distances of each data point to its closest vector $\boldsymbol{\mu}_k$ is minimum.

objective function

$$J = \sum_{i=1}^{N} \sum_{k=1}^{K} r_{ik} \|\mathbf{x}_i - \boldsymbol{\mu}_k\|_2^2$$

the iterative procedure

0. choose initial values for the $\boldsymbol{\mu}_k$

1. minimize J w.r.t. $r_{ik}, \forall i, k$, keeping the $\boldsymbol{\mu}_k$ fixed

    - J involving different $i = 1, 2, \ldots, N$ are independent, so we can optimize for each i separately

        - choose r_{ik} to be 1 for whichever value of k gives the minimum value of $\|\mathbf{x}_i - \boldsymbol{\mu}_k\|_2^2$
        - in other words, we simply assign the i^{th} data point to the closest cluster center
        - $r_{ik} = \begin{cases} 1 & \text{if } k = \arg\min_j \|\mathbf{x}_i - \boldsymbol{\mu}_j\|^2 \\ 0 & \text{otherwise.} \end{cases}$

2. minimize J w.r.t. $\boldsymbol{\mu}_k$, keeping the r_{ik} fixed

    - J is convex w.r.t. $\boldsymbol{\mu}_k, \forall k = 1, 2, \ldots, K$

        - $\frac{\partial J}{\partial \boldsymbol{\mu}_k} = 2 \sum_{i=1}^{N} r_{ik} \left(\mathbf{x}_i - \boldsymbol{\mu}_k\right)$
        - By setting, the gradient to 0, $\boldsymbol{\mu}_k = \frac{\sum_i r_{ik} \mathbf{x}_i}{\sum_i r_{ik}}$

repeat 1, 2 until convergence

Proof of convergence

convergence of the objective function

- $J \geq 0$ because it is a sum of squares

- Every iteration, J is strictly decreasing before the convergence

    - when $\boldsymbol{\mu}_k$'s are fixed, optimize over r_{ik} will reduce the value of J
    - when r_{ik} fixed, optimize over $\boldsymbol{\mu}_k$ will reduce the value of J

- By the monotone convergence theorem, we know J will converge

convergence of the $\boldsymbol{\mu}_k$'s

- Because N and K are limited, the number of the partitions is limited
  - J has only limited number of values
  - $\exists T$, when the number of iterations $> T$, J is a constant $\Rightarrow$ any $\boldsymbol{\mu}_k$ will not change