Name: Yuxuan Zhang yuxuanz8

University of Illinois

Spring 2020

CS 446/ECE 449 Machine Learning

Homework 7: K-Means

Due on Tuesday April 7 2020, noon Central Time

1. **[16 points]** K-Means

We are given a dataset $\mathcal{D} = \{(x)\}$ of 2d points $x \in \mathbb{R}^2$ which we are interested in partitioning into K clusters, each having a cluster center μ_k ($k \in \{1, \ldots, K\}$) via the k-Means algorithm. This algorithm optimizes the following cost function:

$$\min_{\mu_k, r} \sum_{x \in \mathcal{D}, k \in \{1,\ldots,K\}} \frac{1}{2} r_{x,k} \|x - \mu_k\|_2^2 \qquad \text{s.t.} \quad \begin{cases} r_{x,k} \in \{0,1\} & \forall x \in \mathcal{D}, k \in \{1,\ldots,K\} \\ \sum_{k \in \{1,\ldots,K\}} r_{x,k} = 1 & \forall x \in \mathcal{D} \end{cases} \tag{1}$$

(a) (1 point) What is the domain for μ_k?

Your answer: $\mu_k \in R^2$ the same domain as x

(b) (3 points) Given fixed cluster centers μ_k $\forall k \in \{1, \ldots, K\}$, what is the optimal $r_{x,k}$ for the program in Eq. (1)? Provide a reason?

Your answer:

$$r_{x,k} = \begin{cases} 1 & \text{if } k = \arg\min_{k \in \{1,\ldots,K\}} \|x - \mu_k\|_2^2 \\ 0 & \text{otherwise.} \end{cases}$$

The objective function is squared sum and one x ~~the point~~ can only be assigned to one cluster. in order to find optimal $r_{x,k}$. x, μ_k are all fixed. we only need to find the ~~k which~~ whose distance with x is smallest.

(c) (3 points) Given fixed $r_{x,k}$ $\forall x \in \mathcal{D}, k \in \{1, \ldots, K\}$, what are the optimal cluster centers μ_k $\forall k \in \{1, \ldots, K\}$ for the program in Eq. (1)? Reason by first computing the derivative w.r.t. μ_k.

Your answer: $f = \sum_{\substack{x \in D \\ i \in \{1,\ldots,K\}}} \frac{1}{2} r_{x,k} \|x - \mu_k\|^2$

$$\nabla_{\mu_k} f = \sum_{x \in D} r_{x,k} (x - \mu_k) = 0.$$

$$\mu_k = \frac{\sum_{x \in D} r_{x,k} \cdot x}{\sum_{x \in D} r_{x,k}}$$

Name: Yuxuan Zhang yuxuanz8

(d) (5 points) Using Pseudo-code, sketch the algorithm which alternates the aforementioned two steps. Is this algorithm guaranteed to converge? Reason? Is this algorithm guaranteed to find the global optimum? Reason?

Your answer:

① Initialize k centers
② Assign data points to the closest center.
③. Recompute the center of each clusters.
Repeat ② and ③ until they converge. the cluters that data point belong don't change it.

The algorithm is guaranteed to converge. There is finite partition. of D., k^N, so based on our stop criterion. our algorithm will converge in finite step.

The algorithm is not guaranteed. to find the global optimum. because result is depend on the initialization.

(e) (4 points) Complete A7_KMeans.py by implementing the aforementioned two steps. For the given dataset, after how many updates does the algorithm converge, what cost function value does it converge to and what are the obtained cluster centers?

Your answer: 0.5*torch.sum((x-ctmp)**2, axis=1).

After ~~ste~~ two updates the algorithm converges.

cost: ~~9.119990~~. 4.559995

center: [[[1.9163], [-1.9143]], [[-2.0952], [2.0540]]]