B365 Homework 8

Zac Monroe

December 2018

1. (a) $D = \{(0,1)^t, (1,1)^t, (0,2)^t, (1,2)^t, (0,3)^t, (1,3)^t\}$, $m_1 = (0,1)^t$, $m_2 = (0,2)^t$. The initial partition is this: The points that are closest to m_1 are $\{(0,1)^t, (1,1)^t\}$. The mean of these points is $\frac{1}{2}(0+1, 1+1)^t = (\frac{1}{2}, 1)$. The points that are closest to m_2 are $\{(0,2)^t, (1,2)^t, (0,3)^t, (1,3)^t\}$. The mean of these points is $\frac{1}{4}(0+1+0+1, 2+2+3+3)^t = (\frac{1}{2}, \frac{5}{2})^t$.

   (b) $m_1 := (\frac{1}{2}, 1)^t$
   $m_2 := (\frac{1}{2}, \frac{5}{2})^t$
   Closest points to m_1 are $\{(0,1)^t, (1,1)^t\}$. Mean is $\frac{1}{2}(0+1, 1+1)^t = (\frac{1}{2}, 1)$.
   Closest points to m_2 are $\{(0,2)^t, (1,2)^t, (0,3)^t, (1,3)^t\}$ Mean is $\frac{1}{4}(0+1+0+1, 2+2+3+3)^t = (\frac{1}{2}, \frac{5}{2})^t$.

   (c,d) Neither the means nor the partitions changed in the previous step. This indicates that the algorithm is complete, and clusters have been found. Final clusters are $\{(0,1)^t, (1,1)^t\}$ with mean $(\frac{1}{2}, 1)^t$ and $\{(0,2)^t, (1,2)^t, (0,3)^t, (1,3)^t\}$ with mean $(\frac{1}{2}, \frac{5}{2})^t$.

2. $$\begin{aligned} H(m) &= \sum_{i=1}^{n} (x_i - m)^2 \\ &= \sum_{i=1}^{n} (m - x_i)^2 \\ \frac{dH}{dm} &= 2\sum_{i=1}^{n} (m - x_i) \\ &= 2\sum_{i=1}^{n} m - 2\sum_{i=1}^{n} x_i \\ &= 2n \cdot m - 2\sum_{i=1}^{n} x_i. \end{aligned}$$

   Setting $\frac{dH}{dm} = 0$, we have

   $$\begin{aligned} 0 &= 2n \cdot m - 2\sum_{i=1}^{n} x_i \\ 2n \cdot m &= 2\sum_{i=1}^{n} x_i \\ m &= \frac{1}{n}\sum_{i=1}^{n} x_i \end{aligned}$$

3. (a) $H = \sum_{k=1}^{K} \sum_{i:c(i)=k} (x_i - m_k)^2$

   One of the two steps in each iteration of the algorithm is redefining $\{m_k\}$ to be the average of the points $\{x_i\}$ for which $c(i) = k$. See my solution to problem 2 above for a proof that the minimizing value for $\sum_{i=1}^{n} (x_i - m)^2$ is the average of $\{x_i\}$, or $\frac{1}{n}\sum_{i=1}^{n} x_i$. So if m changes, it changes to a value that lowers H. If m does not change, it is already the average of the points $\{x\}$ closest to it, so it is already the minimizing value for H.

   The other step is re-grouping the data points based on the prototype to which each one is closest. This action, by definition, decreases H because the distance between a point x_i and a prototype m_k is $|x_i - m_k|$, which is the square root of $(x_i - m_k)^2$, so those values behave similarly. Choosing the prototype that minimizes the distance to a point therefore reduces the value of $(x_i - m_k)^2$.

(b) The prototypes $\{m_k\}$ only change if there are points that are newly close to them after they were moved to the centroid of their previously-associated points. H is directly dependent on the value of those prototypes, and as shown in my solution to part (a), the value of H monotonically decreases. There are only a finite number of data points $\{x_i\}$, and $\{m_k\}$ can only get so close to each data point, and H never increases, so the prototypes cannot move back to where they were before.

4. By having $m_1 := (0, 2)$ and $m_2 := (1, 2)$, $H = 4$ and the configuration is stable. The configuration from problem 1 resulted in $H = 2\sqrt{2} + 2 \approx 3.8 < 4$, so the configuration depicted below is not globally optimal.