UNIVERSITY OF CALIFORNIA, BERKELEY

HW 7: CS - 189

Oscar Ortega

July 16, 2021

I certify that all solutions are entirely in my own words and that I have not looked at another student's solutions. I have given credit to all external sources I consulted. - Oscar Ortega
I worked on this homework with Aditya Jhanwar

1 REGULARIZED AND KERNEL K-MEANS

- a: If the number of clusters equals the number of sample points. Then we can assign point x_i to be the member of $C_i : \forall i$ WLOG

  This implies the following:

$$\mu_i = \frac{1}{1} x_i = x_i \tag{1.1}$$

$$\min_{C_1,\ldots,C_n} \sum_{i=1}^{n} \sum_{x_j \in C_i} \|x_j - \mu_i\|_2^2 = \sum_{i=1}^{n} \|x_i - x_i\|_2^2 \tag{1.2}$$

$$= 0 \tag{1.3}$$

- b: let $\mathscr{L}(\mu_i) = \left(\lambda \|\mu_i\|_2^2 + \sum_{x_j \in C_i} \|x_j - \mu_i\|_2^2\right)$

$$
\begin{aligned}
\nabla \mathscr{L}_{\mu_i} &= 2\lambda\mu_i + \sum_{x_j \in C_i} \nabla_{\mu_i} \langle x_j - \mu_i, x_j - \mu_i \rangle && (1.4)\\
&= 2\lambda\mu_i + \sum_{x_j \in C_i} \nabla_{\mu_i} \langle x_j, x_j \rangle + \nabla_{\mu_i} - 2\langle x_j, \mu_i \rangle + \nabla_{\mu_i} \langle \mu_i, \mu_i \rangle && (1.5)\\
&= 2\lambda\mu_i + \sum_{x_j \in C_i} -2x_i + 2\mu_i = 0 && (1.6)\\
\lambda\mu_i + |C_i|\mu_i &= \sum_{x_j \in C_i} X_j && (1.7)\\
\mu_i &= \frac{\sum_{x_j \in C_i} X_j}{\lambda + |C_i|} && (1.8)
\end{aligned}
$$

- c:
  - Let μ_i = the location of car i
  - Let C_i = the set of students who decide to take a ride by car i.
  - Let $\mathscr{C} = \{C_1, ..., C_k\}$
  - In this case, we want to minimize the distance the students have to travel as well as the distance the cars have to travel. Our objective will therefore sum up $\|x_j - \mu_i\|_2 : \forall j, i$, the distance each student j has to travel to the closest shuttle.

  Given the set of students, we can now set out to minimize the following objective.

$$
\begin{aligned}
d^* &= \min_{\mathscr{C}} \sum_{i=1}^{k} \sum_{x_j \in C_i} \|x_j - \mu_i\|_2 && (1.9)\\
&= \min_{\mathscr{C}} \sum_{i=1}^{k} \sum_{x_j \in C_i} \|x_j - \mu_i\|_2^2 && (1.10)
\end{aligned}
$$

  Because the domain of our objective function is in $[0, \infty)$, and norms are monotonically increasing functions. Applying a monotonically increasing function such as squaring will return the same set of minimizers.

- d: Consider the K-means objective function

$$\min_{\mathscr{C}} \sum_{i=1}^{k} \varphi(i)$$

  Where I define $\varphi(i) = \sum_{x_j \in C_i} \|x_j - \mu_i\|_2^2 = \sum_{x_j \in C_i} \|x_j - \frac{1}{|C_i|} \sum_{x_k \in C_i} x_k\|_2^2$ Because we want to perform a kernelized version of K-means we now set out to minimize the following kernelized objective function.

$$\min_{\mathscr{C}} \sum_{i=1}^{k} \Phi(\varphi(i))$$

Where $\Phi(\varphi(i)) = \sum_{x_j \in C_i} \|\Phi(x_j) - \Phi(\mu_i)\|_2^2 = \sum_{x_j \in C_i} \|\Phi(x_j) - \frac{1}{|C_i|}\sum_{x_k \in C_i} \Phi(x_k)\|_2^2$ Note how this implies the cost of one point is $\|\Phi(x_j) - \frac{1}{|C_i|}\sum_{x_k \in C_i} \Phi(x_k)\|_2^2$

$$\|\Phi(x_j) - \frac{1}{|C_i|}\sum_{x_k \in C_i} \Phi(x_k)\|_2^2 = \langle \Phi(x_j) - \frac{1}{|C_i|}\sum_{x_k \in C_i} \Phi(x_k), \Phi(x_j) - \frac{1}{|C_i|}\sum_{x_k \in C_i} \Phi(x_k) \rangle \quad (1.11)$$

$$= \langle \Phi(x_j), \Phi(x_j) \rangle - \frac{2}{|C_i|} \langle \Phi(x_i), \sum_{x_k \in C_i} \Phi(x_k) \rangle + \frac{1}{|C_i|^2} \langle \sum_{x_k \in C_i} \Phi(x_k), \sum_{x_k \in C_i} \Phi(x_k) \rangle \quad (1.12)$$

$$= \langle \Phi(x_j), \Phi(x_j) - \frac{2}{|C_i|} \sum_{x_k \in C_i} \langle \Phi(x_j), \Phi(x_k) \rangle + \frac{1}{|C_i|^2} \sum_{x_j \in C_i} \sum_{x_k \in C_i} \langle \Phi(x_j), \Phi(x_k) \rangle \quad (1.13)$$

$$= k(x_j, x_j) - \frac{2}{|C_i|} \sum_{x_k \in C_i} k(x_j, x_k) + \frac{1}{|C_i|^2} \sum_{x_i \in C_i} \sum_{x_k \in C_i} k(x_i, x_k) \quad (1.14)$$

Which means we can set out to choose the following argmin.

$$\text{argmin}_k \Big(k(x_j, x_j) - \frac{2}{|S_k|} \sum_{x_k \in S_k} k(x_j, x_k) + \frac{1}{|S_k|^2} \sum_{x_i \in S_k} \sum_{x_k \in S_k} k(x_i, x_k) \Big)$$

- e:
  Because the $k(x_j, x_j)$ term in our objective function is independent of the cluster we choose, choosing the argmin of the original problem is equivalent to the following minimization.

  $$\text{argmin}_k \Big(- \frac{2}{|S_k|} \sum_{x_k \in S_k} k(x_j, x_k) + \frac{1}{|S_k|^2} \sum_{x_i \in S_k} \sum_{x_k \in S_k} k(x_i, x_k) \Big)$$

  Furthermore, to save computation we can go ahead and pre-compute all the values of the kernel matrix.

2 Low-Rank Approximation

- a: on iPython Notebook
- b: on iPython Notebook
- c: on iPython Notebook
- d: You notice that at about the 30 rank approximation the resolution of the reduced image begins to be noticeably lower than the full rank image. For the sky image, I begin to notice the drop in the resolution of the image at about the 15 rank approximation. The differences must be due to the higher contrast of the images, and more specifically, there are more "directions" of large variance in the face image than in the image of the sky.