ECE M146 Discussion 7 Solution
Introduction to Machine Learning Friday, May 24, 2019
Instructor: Lara Dolecek
TA: Zehui (Alex) Chen, Ruiyi (John) Wu

1. Here is an example where we would want to regularize clusters. Suppose n students are seated for taking an endterm exam in an $\mathbf{R}^2$ Euclidean room. There are K teaching assistants who must collect the answer scripts once the time is up. The TAs need to figure out good locations to position themselves so that the students can walk to the nearest TA and submit their answers. Once the TAs have all the answer sheets, they must return to the front desk located at (0,0) while handling the returned answer sheets carefully. To reduce the possibility of mishaps related to handling of the papers, write down an objective which can be used to minimize the total distance that both students and TAs need to walk to bring the papers to the front desk. Assume that everyone can walk by taking the shortest path between two points.
Solution:
$$J = \sum_{k=1}^{K} \left[\|\mu_k\|_2^2 + \sum_{n=1}^{N} r_{nk} \|x_n - \mu_k\|_2^2 \right].$$

2. We learned that the μ that maximize

$$\sum_{i=1}^{N}(x_i - \mu)^2$$

for $x_i \in \mathbf{R}$ is given by the mean of $\{x_1, \cdots, x_N\}$,i.e., $\mu^* = \frac{1}{N}\sum_{i=1}^{N} x_i$.

Show that the μ that maximize

$$\sum_{i=1}^{N}(x_i - \mu)^0$$

is given by the mode of $\{x_1, \cdots, x_N\}$. What if $x_i \in \mathbf{R}^n$?

Solution: If $\mu \notin \{x_1, \cdots, x_N\}$, then $J = N$. If $\mu \in \{x_1, \cdots, x_N\}$, then $J = N - |\{x_i | x_i = \mu\}|$ because $0^0 = 0$. Then it is clear that letting μ be the mode maximize the objective function.

3. In class, you learned that the direction that maximize the variance of the projection onto a one-dimensional space is the eigenvector that corresponds to the largest eigenvalue of the data covariance matrix $S = \frac{1}{N}X^T X$. Formally, the solution to the following maximization problem

$$\max_{u_1} u_1^T S u_1 \quad \text{subject to } \|u_1\|^2 = 1,$$

is the eigenvector that corresponds to the largest eigenvalue of S.

Suppose u_2 is orthogonal to u_1 and have unit norm. We want to maximizes the variance of the data projected on u_2. Show that the optimal u_2 is defined by the second eigenvectors of the data covariance matrix S that corresponds to the second largest eigenvalues.

Solution: We use a Lagrange multiplier λ_2 to enforce the unit norm constraint $u_2^T u_2 = 1$. We use Lagrange multipliers η_2 to enforce the constraints that u_2 is orthogonal to u_1. The Lagrangian is then:

$$L(u_2, \lambda_2, \eta_2) = u_{M+1}^T S u_{M+1} + \lambda_2(1 - u_2^T u_2) + \eta_2 u_2^T u_1.$$

Setting $\nabla_{u_2} L(u_2, \lambda_2, \eta_2) = 0$, we get

$$0 = 2Su_2 - 2\lambda_2 u_2 + \eta_2 u_1.$$

Left multiplying with u_1^T and using the orthogonality constraints, we see that $\eta_2 = 0$. We therefore obtain

$$Su_2 = \lambda_2 u_2.$$

This shows that the new vector u_2 is an eigenvector of S. Left multiply both sides with u_2 and use the normalization constraint, we have

$$\lambda_2 = u_2^T S u_2.$$

Then the eigenvector should be the one corresponds to the second largest eigenvalue.