Introduction to Machine Learning
Lab 8: Mean-shift and Label Propagation

Hongteng Xu

May 6, 2022

1 Motivation

- Implement the mean-shift algorithm for data clustering
- When some labels are available, observe the diffusion process achieved by mean-shift.
- What if we implement mean-shift and label propagation with fixed kernels?

2 Tasks

Please read Lecture 10 carefully before doing this lab work.

1. **Mean-shift:** Given a set of data $\boldsymbol{X} \in \mathbb{R}^{N\times D}$, implement the mean-shift algorithm to find the means of the clusters, i.e., for the t-th iteration

$$m^{(t)}(\boldsymbol{x}_n) = f_{K_h}(m^{(t-1)}(\boldsymbol{X})) = \sum_{m=1}^{N} \frac{K_h(m^{(t-1)}(\boldsymbol{x}_n), m^{(t-1)}(\boldsymbol{x}_m))m^{(t-1)}(\boldsymbol{x}_m)}{\sum_{i=1}^{N} K_h(m^{(t-1)}(\boldsymbol{x}_n), m^{(t-1)}(\boldsymbol{x}_i))} \tag{1}$$

2. **Label propagation:** When some data points are labeled by one-hot vectors, i.e., $\boldsymbol{Y} \in \{0,1\}^{N\times K}$ and $\boldsymbol{Y}\mathbf{1}_K \leq \mathbf{1}_N$, where K is the number of clusters, we can estimate the labels for the unlabeled data via a **label propagation** algorithm. Essentially, this algorithm treats the observed one-hot vectors as the distribution of labels. It works a variant of the above mean-shift algorithm, in which the kernels are defined based on the mean of samples (i.e., $m(\boldsymbol{x})$'s), while the input and output are the distributions of labels (i.e., normalized $\boldsymbol{y}$'s). (Hint: It can also be viewed as a Nadaraya-Watson estimator)

3. In Task 1 and 2, the kernel matrices are time-varying. What if we use time-invariant kernel $K_h(\boldsymbol{x}_n, \boldsymbol{x}_m)$ instead when implementing the mean-shift and the label-propagation algorithm? Observe their differences and think about the reasons behind the phenomena.