Tsinghua-Berkeley Shenzhen Institute
LEARNING FROM DATA
Fall 2024

## Writing Assignment 3

**Issued:** Friday 15th November, 2024 **Due:** Friday 29th November, 2024

### POLICIES

- **Acknowledgments:** We expect you to make an honest effort to solve the problems individually. As we sometimes reuse problem set questions from previous years, covered by papers and web pages, we expect the students **NOT** to copy, refer to, or look at the solutions in preparing their answers (relating to an unauthorized material is considered a violation of the honor principle). Similarly, we expect you to not google directly for answers (though you are free to google for knowledge about the topic). If you do happen to use other material, it must be acknowledged here, with a citation on the submitted solution.
- **Required homework submission format:** You can submit homework either as one single PDF document or as handwritten papers. Written homework needs to be provided during the class on the due date, and a PDF document needs to be submitted through Tsinghua's Web Learning (`http://learn.tsinghua.edu.cn/`) before the end of the due date.
- **Collaborators:** In a separate section (before your answers), list the names of all people you collaborated with and for which question(s). If you did the HW entirely on your own, **PLEASE STATE THIS**. Each student must understand, write, and hand in answers of their own.

3.1. (2 points) Let $X = \mathbb{R}$ and let $C = \{c_{a,b} \mid a, b \in \mathbb{R}, a < b\}$ be the concept class of intervals, where $c_{a,b} : \mathbb{R} \to \{0, 1\}$ is defined as $c_{a,b}(x) = 1$ if $x \in [a, b]$ and 0 otherwise. Please show that $\mathrm{VCdim}(C) = 2$.

3.2. (Bonus 2 points) (Rademacher Complexity)

(a) (1 point) Given any function class $\mathcal{F}$ and constants $a, b \in \mathbb{R}$, denote the function class

$$\{g|g(x) = af(x) + b\}$$

by $a\mathcal{F} + b$. Show that the Rademacher complexity of $a\mathcal{F} + b$ is:

$$R_m(a\mathcal{F} + b) = |a|R_m(\mathcal{F})$$

(b) (1 point) Given a class $\mathcal{H}$ of binary classifiers $\mathcal{H} = \{h : \mathcal{X} \to \mathcal{Y}\}$ where $\mathcal{Y} = \{-1, +1\}$, we can define a class of 0-1 loss functions $l_h : \mathcal{X} \oplus Y \to \mathbb{R}$ as follows:

$$L(\mathcal{H}) = \{l_h \mid l_h(x, y) = \frac{1 - h(x)y}{2}, x \in \mathcal{X}, y \in \mathcal{Y}, h \in \mathcal{H}\}$$

Use the previous result to show that $2R_m(\mathcal{H}) = R_m(L(\mathcal{H}))$.

3.3. (2+1 points) (K-means) Given input data $\mathcal{X} = \{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(m)}\}$, $\boldsymbol{x}^{(i)} \in \mathbb{R}^d$, the $k$-means clustering partitions the input into $k$ sets $C_1, \ldots, C_k$ to minimize the within-cluster sum of squares:

$$\arg\min_{C} \sum_{j=1}^{k} \sum_{\boldsymbol{x} \in C_j} \|\boldsymbol{x} - \boldsymbol{\mu}_j\|^2,$$

where $\boldsymbol{\mu}_j$ is the center of the $j$-th cluster:

$$\boldsymbol{\mu}_j \overset{\text{def}}{=} \frac{1}{|C_j|} \sum_{\boldsymbol{x} \in C_j} \boldsymbol{x}, \quad j = 1, \ldots, k.$$

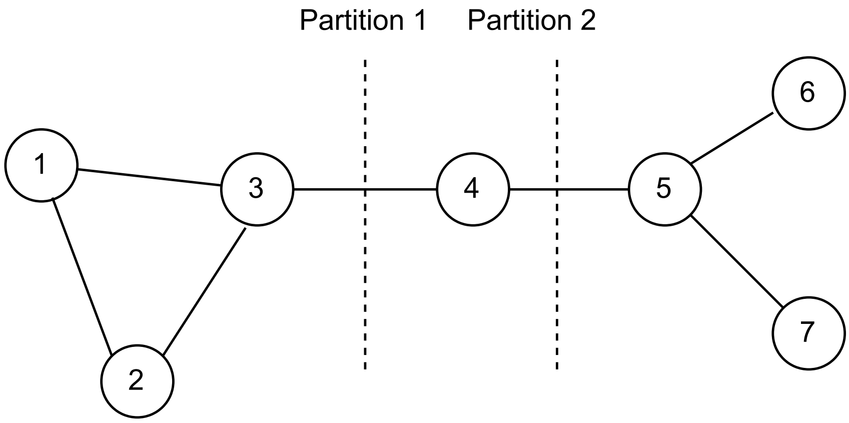
(a) (2 points) Show that the $k$-means clustering problem is equivalent to minimizing the pairwise squared deviation between points in the same cluster:

$$\sum_{j=1}^{k} \frac{1}{2|C_j|} \sum_{\boldsymbol{x}, \boldsymbol{x}' \in C_j} \|\boldsymbol{x} - \boldsymbol{x}'\|^2.$$

(b) (bonus 1 points) Show that the $k$-means clustering problem is equivalent to maximizing the between-cluster sum of squares:

$$\sum_{i=1}^{k} \sum_{j=1}^{k} |C_i||C_j|\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|^2.$$

3.4. (2 points) (Spectral Clustering) Consider the following unweighted graph.



(a) (1 point) Compute the *RatioCut* values for the binary partition of the graph at edge (3,4) and (4,5), respectively. Which partition is better? Repeat the same analysis for *NormalizedCut* (NCut). Discuss the differences between the two balanced graph cut approaches.

(b) (1 point) Perform the spectral clustering algorithm and show the steps. e.g. Show the normalized Laplacian and its eigenvectors, then give the cluster label. (You may use a computer to perform the eigen decomposition.)