



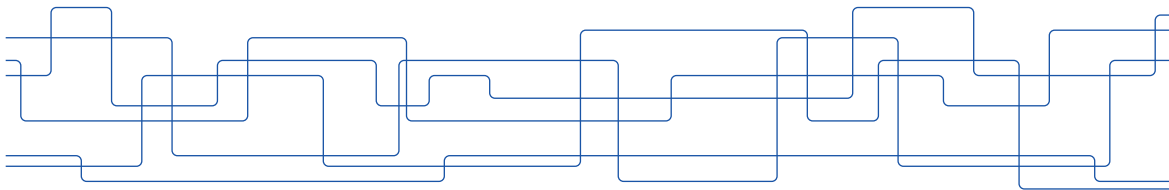
DD2434 Machine Learning, Advanced Course

Exercise session on modules 8 and 9

Aristides Gionis

argioni@kth.se

KTH Royal Institute of Technology



question 1. (averaging reduces variance)

- ▶ let X be a random variable with $\mathbb{E}[X] = \mu$ and $\text{Var}[X] = \sigma^2 < \infty$
let $Y = \frac{1}{k} \sum_{i=1}^k X_i$, where $\{X_i\}_{i=1,\dots,k}$ are k independent copies of X
- ▶ show that $\mathbb{E}[Y] = \mu$ and $\text{Var}[Y] = \frac{\sigma^2}{k}$

question 2.1. (boosting the probability of success)

- ▶ let X be a random variable that estimates a quantity of interest μ , i.e.,

$$\mathbb{E}[X] = \mu \tag{1}$$

- ▶ consider also the following assertion, for some $\epsilon > 0$:

$$\mathbb{P}[(1 - \epsilon)\mu \leq X \leq (1 + \epsilon)\mu] \geq 3/4 \tag{2}$$

- ▶ does (1) always imply (2), for any $\epsilon > 0$?

question 2.2. (boosting the probability of success)

- ▶ assume that the following holds, for some $\epsilon > 0$:

$$\mathbb{P}[(1 - \epsilon)\mu \leq X \leq (1 + \epsilon)\mu] \geq 3/4$$

- ▶ how can we **boost** the probability of success from $3/4$ to $1 - \delta$, for any $\delta > 0$?
in other words, we want

$$\mathbb{P}[(1 - \epsilon)\mu \leq Z \leq (1 + \epsilon)\mu] \geq 1 - \delta$$

for some random variable Z associated with X

question 2.3. (boosting the probability of success)

(quantitative part of Q2.2)

- how large should k be for

$$\mathbb{P}[(1 - \epsilon)\mu \leq Z \leq (1 + \epsilon)\mu] \geq 1 - \delta$$

to hold?

question 3.1. (graph Laplacian)

- ▶ let $G = (V, E)$ be an undirected d -regular graph
let \mathbf{A} be the adjacency matrix of G
and let $\mathbf{L} = I - \frac{1}{d}\mathbf{A}$ be the **normalized Laplacian** of G
- ▶ show that there is a matrix \mathbf{B} so that $\mathbf{L} = \mathbf{B}^T \mathbf{B}$

question 4.1 (ratio cut)

- for a graph $G = (V, E)$ and partition $S_1, S_2 \subseteq V$, $S_1 \cup S_2 = V$, $S_1 \cap S_2 = \emptyset$
we define the **ratio cut**

$$\text{RatioCut}(S_1, S_2) = \sum_{i=1}^2 \frac{|(u, v) \in E : u \in S_i \text{ and } u \notin S_i|}{|S_i|} = \sum_{i=1}^2 \frac{\text{Cut}(S_i, \bar{S}_i)}{|S_i|}$$

define vector $\mathbf{a} \in \mathbb{R}^n$, so that

$$a_u = \begin{cases} \sqrt{|\bar{S}|/|S|} & \text{if } u \in S \\ -\sqrt{|S|/|\bar{S}|} & \text{if } u \in \bar{S} \end{cases}$$

show that

$$\mathbf{a}^T \mathbf{L} \mathbf{a} = \frac{n}{d} \text{RatioCut}(S, \bar{S})$$

question 4.2 (ratio cut)

- ▶ what are the properties of this particular definition of the **a** vector that make it **useful** in this context of graph partitioning?

recall:

$$a_u = \begin{cases} \sqrt{|\bar{S}|/|S|} & \text{if } u \in S \\ -\sqrt{|S|/|\bar{S}|} & \text{if } u \in \bar{S} \end{cases}$$

question 4.3. (ratio cut)

- ▶ what is the implication that there is a particular vector \mathbf{a} for which

$$\mathbf{a}^T \mathbf{L} \mathbf{a} = \frac{n}{d} \text{RatioCut}(S, \bar{S})$$

and in addition, $\mathbf{a} \neq \mathbf{0}$ and $\sum_u a_u = 0$

- ▶ in particular, how do use this information to design a method for finding a graph partitioning that optimizes RatioCut ?

question 5. (why graph neural networks?)

- ▶ graph neural networks are inspired by neural networks and deep learning
 - deep learning has been very successful in image recognition
 - a graph is represented by its adjacency matrix
 - a vertex is represented by a row-vector in the adjacency matrix
- ▶ why can't we view the adjacency matrix of a graph as an image, and apply "standard" deep learning methods?

question 6. (graph neural networks)

- ▶ how can we implement graph neural networks for graphs with no vertex attributes?