Mathematical Foundations of Data Science - Winter 2022-2023

Practice Final Exam

Tuesday 31.01.2023

Name: (Print Clearly) Sam's Solutions + grading key.

Question 1	12	
Question 2	15	
Question 3	12	
Question 4	11	
Total	50	

- There are 4 questions.
- You must show your work on all problems. Make clear your order of thinking and cross out any work that you don't want to be graded.
- If at any time you struggle with english words or phrasing on the exam you are welcome to use German, I need the practice anyways.
- You have 90 minutes on the exam. Raise your hand if you have a question.

Night Cafe's AI generated "Mathematical foundations of data analysis" Good luck!

1. Recall: **Lemma 1** The map L induced by the Laplacian of a graph $G = (V, E)$ is given by

$$Lf(u) = \frac{1}{\sqrt{\deg(u)}} \sum_{v \in V : \{u,v\} \in E} \frac{f(u)}{\sqrt{\deg(u)}} - \frac{f(v)}{\sqrt{\deg(v)}}.$$

Need non-isolated vertices!

(a) (6 points) Construct a nonzero function f so that $Lf(v) = 0$ for all $v \in V$. This shows that $\lambda = 0$ is always an eigenvalue of the Laplacian.

(b) (6 points) Let G be a *complete* bipartite graph with bipartition $V_1 \sqcup V_2$. Suppose there is a nonzero function $\tilde{f}$ so that $L\tilde{f} = \tilde{f}$. ~~Show that 2 is an eigenvalue of L.~~ Construct lin. ind. function $\hat{f}$ so $L\hat{f} = \hat{f}$.

(a) Define $f(u) = \sqrt{\deg(u)}$, +2

If $u \in V_1$, $Lf(u) = \sqrt{\deg(u)} - \sum_{v \in V_2} \frac{\sqrt{\deg(v)}}{\sqrt{\deg(u)\deg(v)}}$

complete $\Rightarrow$ $\sum_{\substack{u \in V \\ \{u,v\} \in E}} = \sum_{v \in V_2}$ +2

$$= \sqrt{\deg(u)} - \frac{1}{\sqrt{\deg(u)}} \boxed{\sum_{v \in V_2} 1}$$

degree of u! +2

$$= \sqrt{\deg(u)} - \sqrt{\deg(u)} = 0.$$

Similar when $u \in V_2$.

(b) Define $\hat{f}(v) = \begin{cases} \tilde{f}(v) & v \in V_1 \\ -\tilde{f}(v) & v \in V_2 \end{cases}$ +2

Then for $u \in V_1$

$$L\hat{f}(u) = \hat{f}(u) - \sum_{v \in V_2} \frac{\hat{f}(v)}{\sqrt{\deg(u)\deg(v)}}$$

$$= \tilde{f}(u) + \sum_{v \in V_2} \frac{\tilde{f}(v)}{\sqrt{\deg(u)\deg(v)}}$$

By Lemma 7, $\widetilde{L} f(u) = \tilde{f}(u) - \sum \frac{\tilde{f}(v)}{\sqrt{\deg(u)\deg(v)}}$

+2

$$\Rightarrow \sum \frac{\tilde{f}(v)}{\sqrt{\deg(u)\deg(v)}} = \tilde{f}(u) - L\tilde{f}(u)$$

$$L\hat{f}(u) = \tilde{f}(u) + \tilde{f}(u) - L\tilde{f}(u)$$

$$= 2\tilde{f}(u) - \tilde{f}(u)$$

$$= \tilde{f}(u).$$

+4

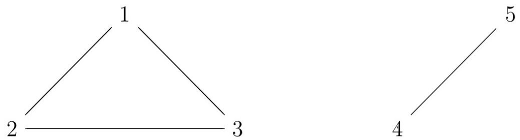
Similar when $v \in V_2$ with negative sign.

Lastly $\vec{f}^1$ not a multiple of $\vec{f}^2$,

So linearly independent. +1

2. Recall an irreducible Markov process on G with transition matrix P satisfies for all $u, v \in V$ there exists $k \in \mathbb{N}$ with $(P^k)_{uv} > 0$. Also recall a unique stationary distribution $\pi : V \to \mathbb{R}$ satisfies $P\pi = \pi$.

   Now consider the following graph:



   with the uniform Markov process. $\rightarrow \frac{1}{\deg(v)}$

   (a) (5 points) Compute the transition matrix P of X.

   (b) (5 points) Show X cannot be irreducible.

   (c) (5 points) Give a stationary distribution $\pi : V \to \mathbb{R}$

(a)
$$P = \begin{array}{c|ccccc} & 1 & 2 & 3 & 4 & 5 \\ \hline 1 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 2 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 3 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ 4 & 0 & 0 & 0 & 0 & 1 \\ 5 & 0 & 0 & 0 & 1 & 0 \end{array}$$

+1 point for each row

(b) Recall $(P^k)_{uv}$ is the probability of travelling from u to v in k steps. +2

The probability of going from 1 to 5 is always 0 since they are not connected by an edge. +2

Thus for all $k \in \mathbb{N}$

$$(P^k)_{15} = 0.$$ +1

$\Rightarrow$ Not irreducible.

(c) By symmetry $\pi(1) = \pi(2) = \pi(3)$

$\pi(4) = \pi(5)$

I'll try $\pi(1) = \pi(2) = \pi(3) = \pi(4) = \pi(5) = \frac{1}{5}.$

$$P\begin{pmatrix} 1/5 \\ \vdots \\ 1/5 \end{pmatrix} = \left(\begin{bmatrix} 0 & 1/2 & 1/2 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}\right)\begin{bmatrix} 1/5 \\ 1/5 \\ 1/5 \\ 1/5 \\ 1/5 \end{bmatrix}$$

Note it is not unique here!

$$= \begin{bmatrix} 2/10 \\ 2/10 \\ 2/10 \\ 1/5 \\ 1/5 \end{bmatrix} = \begin{bmatrix} 1/5 \\ 1/5 \\ 1/5 \\ 1/5 \\ 1/5 \end{bmatrix}$$

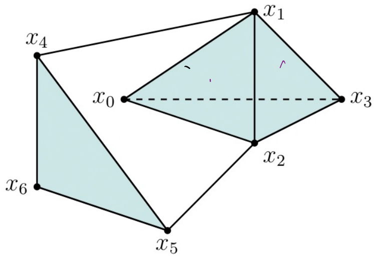
+3 points finding π

+2 points justifying

(c) (5 points) Give a stationary distribution $\pi : V \to \mathbb{R}$

3. Consider the following simplicial complex K. Assume that $\{0, 1, 2, 3\}$ is included in K.

??



(a) (6 points) Compute the Euler Characteristic of K, by both methods.

i. $\chi(K) = \sum_{i \geq 0} (-1)^i k_i$ where k_i is the number of i-dimensional simplices.

ii. $\chi(K) = \sum_{i \geq 0} (-1)^i \beta_i(K)$ where $\beta_i(K)$ which is the number of connected components for $i = 0$, and the number of n-dimensional holes for $i \geq 1$.

(i) k_0 = # vertices = 7

k_1 = # edges = 11

k_2 = # triangles = 5

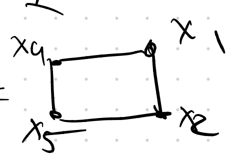
k_3 = # tetrahedron = 1

+1

$\chi(K) = \underbrace{7 - 11 + 5 - 1}_{\text{(+1) alternating}} = 0$ +1

(ii) $\beta_0(k)$ = # connected components = 1

$\beta_1(K)$ = # 1-dim'l holes = 1 =



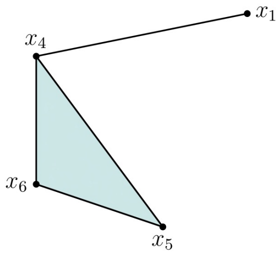
+1

$\beta_2(K)$ = # 2-dim'l holes = 0 → tetrahedron included so no hole

+1

$\chi(K) = 1 - 1 + 0 = 0.$ +1

(b) (6 points) Consider the following subset of K:



Compute $C_1(K)$, $C_2(K)$, $\ker(\partial_2)$ and $\operatorname{Im}(\partial_2)$ where $\partial_2 : C_2(K) \to C_1(K)$ is the boundary operator.

+1

$C_1(K) = \text{Span}\{\{x_1, x_4\}, \{x_4, x_5\}, \{x_5, x_6\}, \{x_4, x_5\}\}$

$C_2(K) = \text{Span}\{\{x_4, x_5, x_6\}\}.$ +1

$\partial_2 : C_2(K) \to C_1(K)$

$\partial_2(\{x_4, x_5, x_6\}) = \{x_4, x_5\} + \{x_5, x_6\} + \{x_4, x_6\}$

$\ker(\partial_2) = \{0\}$ $\quad$ dim $= 0$ +2

$\operatorname{im}(\partial_2) = \text{Span}\{\{x_4, x_5\} + \{x_5, x_6\} + \{x_4, x_6\}\}$ dim $= 1$

+2

operator.

4. Recall in support vector machines given data $\{x_1, \ldots, x_n\} \subseteq \mathbb{R}^D$, we categorize data into $\{y_1, \ldots, y_n\} \subseteq \{\pm 1\}$ by solving an optimization problem, such as the Soft Margin SVM:

$$\min_{a,b,\xi} \|b\|^2 + C \sum_{k=1}^{n} \xi_k$$

subject to the constraints

$$y_k(a + \langle b, x_k \rangle) \geq 1 - \xi_k \text{ for } \xi_k \geq 0, k - 1, \ldots, n.$$

for $a \in \mathbb{R}$, $b \in \mathbb{R}^D$, $x_1, \ldots, x_n \in \mathbb{R}^D$.

(a) (6 pts) The Karush–Kuhn–Tucker conditions (KKT) say that the Soft Margin SVM is solved when

$$\frac{\partial \mathcal{L}}{\partial a} = 0, \quad \frac{\partial \mathcal{L}}{\partial b} = 0, \quad \frac{\partial \mathcal{L}}{\partial \xi} = 0,$$

where for $\alpha_k, \beta_k \geq 0$

$$\mathcal{L}(a, b, \xi, \alpha, \beta) = \|b\|^2 + C \sum_{k=1}^{n} \xi_k - \sum_{k=1}^{n} \alpha_k((y_k(a + \langle b, x_k \rangle) - (1 - \xi_k)) - \sum_{k=1}^{n} \beta_k \xi_k.$$

Show that $b^* = \frac{1}{2} \sum_{k=1}^{n} \alpha_k x_k y_k$ gives the optimal value for Soft Margin SVM.

(b) (5 points) Describe issues with the Soft Margin SVM with respect the the data $x_1 = (1, 1)$, $x_2 = (1, -1)$, $x_3 = (-1, -1)$, $x_4 = (-1, 1)$ and $y_1 = y_3 = -1$ and $y_2 = y_4 = 1$. What is the key to overcoming this problem?

(a) $$\frac{\partial \mathcal{L}}{\partial a} = -\sum_{k=1}^{n} \alpha_k y_k = 0$$

$$\frac{\partial \mathcal{L}}{\partial b} = \left[\frac{\partial \mathcal{L}}{\partial b_i} \right]_{i=1}^{D}$$ +2

$$\frac{\partial \mathcal{L}}{\partial b_i} = 2b_i - \sum_{k=1}^{n} \alpha_k y_k \left(x_k^{(i)} \right)$$

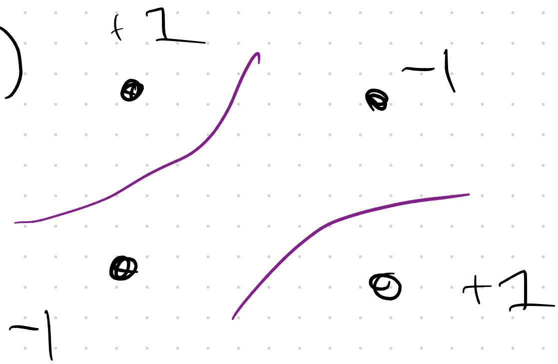
+2

$$\frac{\partial \mathcal{L}}{\partial b} = 2b - \sum_{k=1}^{n} \alpha_k y_k x_k = 0$$

$$\boxed{b = \frac{1}{2} \sum_{k=1}^{n} \alpha_k y_k x_k}$$ +2

this is optimal by KKT

(5)



+3 • Cannot be separated by a plane

+2 • Use kernels to introduce non linear separation problems