Mathematical Foundations of Data Science - Winter 2022-2023
Practice Final Exam
Tuesday 31.01.2023

Name: (Print Clearly)______________________________

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

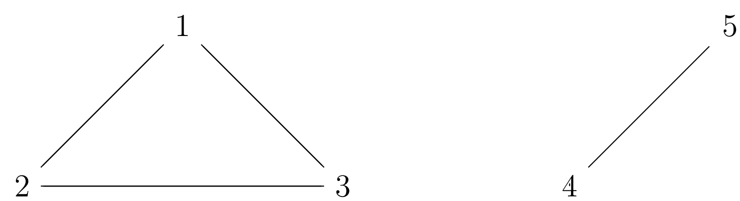
$$Lf(u) = \frac{1}{\sqrt{\deg(u)}} \sum_{v \in V : \{u,v\} \in E} \frac{f(u)}{\sqrt{\deg(u)}} - \frac{f(v)}{\sqrt{\deg(v)}}.$$

   (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.

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:



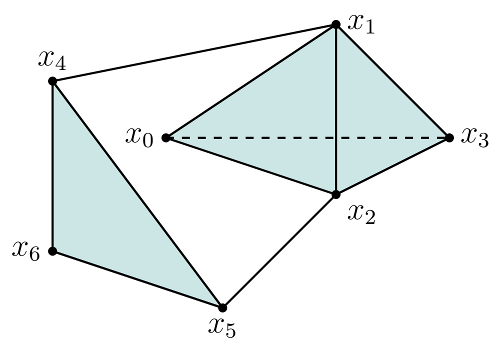
   with the uniform Markov process.

   (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 $\pi : V \to \mathbb{R}$

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

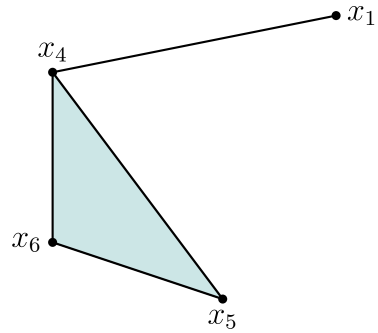


   (a) (6 points) Compute the Euler Characteristic of K, by both methods.

      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 $i \geq 1$.

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



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

4. Recall in support vector machines given data $\{x_1, \dots, x_n\} \subseteq \mathbb{R}^D$, we categorize data into $\{y_1, \dots, y_n\} \subseteq \{\pm 1\}$ 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

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

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

(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 $b^* = \sum_{k=1}^{n} \alpha_k x_k y_k$ 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?