CSE 521: Design and Analysis of Algorithms **Fall 2022**

Problem Set 4

Deadline: Nov 25th in Canvas

1) In this problem we better understand PSD matrices.

   a) Prove or disprove: If $A \in \mathbb{R}^{n\times n}$ is a PSD matrix then $A_{i,i} \geq 0$ for all $1 \leq i \leq n$.

   b) Prove or disprove: If $A \in \mathbb{R}^{n\times n}$ is a PSD matrix, ten $A_{i,j} \geq 0$ for all $1 \leq i, j \leq n$.

   c) For a vector $u \in \mathbb{R}^n$, we write $u \otimes u$ to denote the vector in $\mathbb{R}^{n^2}$ where for any $1 \leq i, j \leq n$, $(u \otimes u)_{in+j} = u_i \cdot u_j$. Show that for any pair of vectors $u, v \in \mathbb{R}^n$,

   $$\langle u \otimes u, v \otimes v\rangle = \langle u, v\rangle^2.$$

   d) Let $A \in \mathbb{R}^{n\times n}$ be a PSD matrix, and let $B \in \mathbb{R}^{n\times n}$ be the matrix where $B_{i,j} = A^2_{i,j}$. Prove that B is PSD.

   Hint: Use part (a) and that any matrix A is PSD iff it can be written as $A = CC^\intercal$ for some matrix $C \in \mathbb{R}^{n\times k}$, for some integer k.

   e) **Extra Credit:** Let $P = \{p_1, \dots, p_n\} \subseteq \mathbb{R}^d$ be a set of points of norm 1. For $\sigma > 0$, let $G_\sigma \in \mathbb{R}^{n\times n}$ be the Gaussian kernel on P, i.e.,

   $$G_\sigma(i, j) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\|p_i - p_j\|^2/2\sigma}.$$

   Prove that $G_\sigma \succeq 0$.

2) Let $A \in \mathbb{R}^{m\times n}$ and suppose we want to estimate Av for a vector $v \in \mathbb{R}^n$. Here is a randomized algorithm for this task. Choose the i-th column of A, A_i, with probability

   $$p_i = \frac{\|A_i\|^2}{\|A\|_F^2}$$

   and let $X = A_i v_i/p_i$. Show that $\mathbb{E}[X] = Av$. Calculate $\mathrm{Var}(X) = \mathbb{E}\left[\|X\|^2\right] - \|\mathbb{E}X\|^2$. Suppose, in preprocessing, we have calculated $p_1, \dots, p_n$ and we have made an array $b[]$ where $b[i] = p_1 + \cdots + p_i$ for all $1 \leq i \leq n$. This should help you to generate a random column in time $O(\log n)$. Design an algorithm that for a given $\epsilon > 0$ and $v \in \mathbb{R}^n$ runs in time $O(m\frac{\log(n)\|A\|_F^2}{\epsilon^2\|A\|^2})$ and outputs a vector $y \in \mathbb{R}^m$ such that with probability at least 9/10

   $$\|Av - y\| \leq \epsilon\|A\|\|v\|$$

   Note that if A has a few large eigenvalues and many small eigenvalues, then we can estimate Av in linear time as opposed to quadratic time using your algorithm.

3) In this part implement to following heuristic to find the hidden partition problem: You are given a graph $G = (V, E)$ with adjacency matrix A with $n = 400$ vertices. The graph is stored in the file "hidden1.in". Let D be a degree matrix where $D_{v,v} = d(v)$ is the degree of a vertex v. Then $\tilde{A} := D^{-1/2}AD^{-1/2}$ is called the normalized adjacency matrix of G. Let x be the second largest eigenvector of $\tilde{A}$; find the median of values in x and output vertices below the median as one community and the rest as the other. This graph is constructed with $p = 0.65$ and $q = 0.05$. The true partition is in the file "hidden1.out". Compare your output with the true hidden partition and report how many misplaced vertices are there in your output.