CS6550: Continuous Algorithms: Optimization and Sampling **Spring 2020**

Homework 1

Lecturer: Santosh Vempala *Due Date: 22 Sept 2021*

Notes:

- You can discuss and collaborate, but please write your own solutions, and clearly mention everyone you discussed with.
- Start on a new page for each problem.
- Submit on Canvas via Gradescope

1. [Mixtures of Convex Bodies]

Let F be a mixture of uniform distributions on k unknown convex bodies $F_1, F_2, \dots F_k$ with mixing weights $w_i \geq 0, \sum_{i=1}^k w_i = 1$. Suppose that each F_i is weakly isotropic, i.e., $\mathbb{E}_{F_i}(x) = \mu_i, \mathbb{E}_{F_i}((x-\mu_i)(x-\mu_i)^\top) = \sigma_i^2 I$.

- Give a pairwise separation condition on the means of the component distributions which would suffice to cluster a sample of m points (according to their component of origin) with high probability.
- Give an efficient algorithm to cluster points according to your condition, and to estimate the mean and covariance matrix of each component.
- Suppose each component is not isotropic, but has a covariance matrix Σ_i satisfying $cI \preceq \Sigma_i \preceq CI$. How would you modify your condition to solve this more general problem?

2. [Variance Separation]

Let $F_1 = N(0, \sigma_1^2 I)$ and $F_2 = N(0, \sigma_2^2 I)$ are two spherical Gaussians in $\mathbb{R}^d$.

- Give a condition on σ_1, σ_2 so that F_1, F_2 are probabilistically separated.
- Suppose $F = \frac{1}{2}F_1 + \frac{1}{2}F_2$ is a mixture of two Gaussians with unknown σ_1, σ_2. Give an algorithm to cluster the sample WHP according to the component distributions under your condition. (For a simpler setting, assume that half of points are drawn from F_1.)

3. [Tensor Maxima]

Let $T \in \mathbb{R}^{n\times n\times n}$ be a symmetric third-order tensor so that T_{ijk} remains the same for any permutation of i, j, k. It defines a polynomial $T(x,x,x) = \sum_{i,j,k} T_{ijk} x_i x_j x_k$ where $x \in \mathbb{R}^n$. The operator norm (or 2-norm) of T is defined as

$$\sup_x f(x) = \frac{T(x,x,x)}{\|x\|_2^3}.$$

- Characterize the local minima/maxima of f.
- How many local minima/maxima can f have?
- Compare this with the case of matrices. What additional properties do local minima/maxima have when T is a matrix?
- Suppose that $T = \sum_{\ell=1}^k \alpha_i u_i \otimes u_i \otimes u_i$ where $\alpha_i \in R$ and $\{u_i\}$ are orthonormal vectors. What are the fixed points of the following tensor power iteration?

$$x \leftarrow \frac{T(x,x,.)}{\|T(x,x,.)\|_2}.$$

4. [Robust Mean]

Suppose we are given ϵ-corrupted samples from a Gaussian distribution $N(\mu, 1)$ with unknown mean $\mu \in \mathbb{R}$. In other words, $(1-\epsilon)$ of the sample is from the Gaussian, and ϵ fraction of the sample is arbitrary, possibly chosen by an adversary with knowledge of the Gaussian.

- Show that the median of the sample ν satisfies $\|\nu - \mu\|_2 = O(\epsilon)$ when the number of samples is $\Omega(1/\epsilon^2)$.
- Now consider the same problem in $\mathbb{R}^d$, where the underlying Gaussian is $N(\mu, I)$ and ϵ fraction of the sample can be placed arbitrarily in $\mathbb{R}^d$. Suppose we define the median as the point whose coordinate in each axis are the median along that axis, i.e., $\nu = (\nu(x_1), \ldots, \nu(x_d))^\top$. How large can $\|\nu - \mu\|_2$ be? You can choose the corruptions to make this error as large as possible.
- The geometric median of a set of points is a point that minimizes the sum of Euclidean distances from the point set to itself. Suppose we use the geometric median as a robust mean estimate in $\mathbb{R}^d$ for the same ϵ-corrupted noisy distribution as above. How large can the error be?