University of Edinburgh
INFR11156: Algorithmic Foundations of Data Science (2019)
Homework 2

Problem 1: Suppose you want to estimate the unkown centre of the Gaussian in $\mathbb{R}^d$ which has unit variance in each direction. Show that $O(\log d/\epsilon^2)$ random samples form the Gaussian are sufficient to get an estimate $\mathbf{m}_s$ of the true centre $\boldsymbol{\mu}$, so that with probability at least 99%,

$$\|\boldsymbol{\mu} - \mathbf{m}_s\|_\infty \leq \epsilon.$$

How many samples are sufficient to ensure that with probability at least 99%

$$\|\boldsymbol{\mu} - \mathbf{m}_s\|_2 \leq \epsilon?$$

Note that $\|\mathbf{x}\|_\infty := \max_i |x_i|$.

Problem 2 (challenging): This question is to try to design a dimension reduction lemma for ℓ_1, similar to the Johnson-Lindenstrauss (JL) lemma for the Euclidean space. Remember that JL lemma says that we can pick a matrix Φ, of dimension $k \times d$ for large enough k, where each entry is chosen from a Gaussian distribution, such that: for any $x \in \mathbb{R}^d$, we have that $\frac{1}{\sqrt{k}} \|\Phi x\|_2$ is a $(1+\epsilon)$ approximation to $\|x\|_2$ with probability at least $1 - \mathrm{e}^{-\Omega(\epsilon^2 k)}$.

For ℓ_1, the equivalent of Gaussian distribution is the Cauchy distribution, which has probability distribution function $p(x) = \frac{1}{\pi(1+x^2)}$. Namely, the corresponding "stability" property of Cauchy distribution is the following. Consider $s = \sum_{i=1}^d x_i c_i$, for $x \in \mathbb{R}^d$ and c_i each independently chosen from Cauchy distribution. Then s has distribution $\|x\|_1 \cdot c$ where c is also distributed as a Cauchy distribution.

It is tempting to construct a dimensionality reducing map for ℓ_1 in the same way as what we did for Euclidean space, just by replacing the Gaussian distribution with the Cauchy distribution. In particular, let C be a matrix of size $k \times d$, where each entry is chosen independently from the Cauchy distribution.

1. Argue that this approach does *not* work for dimensionality reduction for ℓ_1. Namely, for (say) $k = 1000$ and $x = (1, 0, 0, \dots, 0)$, the estimator $\frac{1}{k}\|Cx\|_1$ is not a 2-approximation to $\|x\|_1 = 1$ with probability at least 10%.

   In fact, it has been proven that there does *not* exist an equivalent dimensionality reduction for ℓ_1 at all. Instead, we will construct a sketch that provides a weaker form of "dimension reduction".

2. The *median estimate* is defined as the *median* of the absolute values of k coordinates of the vector Cx. Prove that for any $x \in \mathbb{R}^d$, the median estimate is a $1+\epsilon$ approximation to $\|x\|_1$ with at least $1 - \mathrm{e}^{-\Omega(\epsilon^2 k)}$ probability. You might want to use the following concentration bound, called *Chernoff bound*: for any k independent and identically distributed random variables $x_1, \dots, x_k \in \{0, 1\}$, each with expectation $\mathbf{E}[x_i] = \mu \in [0, 1]$, we have that

$$\mathbf{P}\left[\left|\frac{1}{k}\sum_i x_i - \mu\right| > \epsilon\right] \leq \mathrm{e}^{-\epsilon^2 k}.$$

   Note that we obtain a "sketch", which is not a regular dimension reduction scheme: namely, the "target" space is not ℓ_1, but "median" (which is not even a metric/distance). Nevertheless, it is a linear map, and is useful for sketching and streaming as we will see in future lectures.