University of Edinburgh

INFR11156: Algorithmic Foundations of Data Science (2019)

Homework 2

Solutions

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

Solution: We assume that there are k samples denoted by $\mathbf{x}^1, \ldots, \mathbf{x}^k$, where every $\mathbf{x}^i \in \mathbb{R}^d$ is expressed by $(\mathbf{x}^i_1, \ldots, \mathbf{x}^i_d)$. Notice that the condition $\|\boldsymbol{\mu} - \mathbf{m}_s\|_\infty \leq \epsilon$ holds if and only if $|\boldsymbol{\mu}_i - (\mathbf{m}_s)_i| \leq \epsilon$ for any $1 \leq i \leq d$. Hence, our goal is find that a bound on k such that it holds with probability at least $1 - O(1/d)$ that

$$|\boldsymbol{\mu}_i - (\mathbf{m}_s)_i| \leq \epsilon,$$

for any fixed i. Then applying the union bound gives us the desired statement.

Now we fix an arbitrary i, and without loss of generality we assume that $\boldsymbol{\mu}_i = 1$. We look at the probability that $(\mathbf{m}_s)_i > (1+\epsilon)$. By definition, we have

$$(\mathbf{m}_s)_i = \frac{1}{k}\left(\mathbf{x}^1_i + \cdots + \mathbf{x}^k_i\right),$$

and therefore it holds for any $\lambda > 0$ that

$$\begin{aligned}
\mathbf{P}\left[\,(\mathbf{m}_s)_i > (1+\epsilon)\,\right] &= \mathbf{P}\left[\frac{1}{k}\left(\mathbf{x}^1_i + \cdots + \mathbf{x}^k_i\right) \geq (1+\epsilon)\right] \\
&= \mathbf{P}\left[\mathrm{e}^{\lambda\cdot\sum_{j=1}^k \mathbf{x}^j_i} \geq \mathrm{e}^{(1+\epsilon)\cdot\lambda k}\right] \\
&\leq \frac{\mathbf{E}\left[\mathrm{e}^{\lambda\cdot\sum_{j=1}^k \mathbf{x}^j_i}\right]}{\mathrm{e}^{(1+\epsilon)\cdot\lambda k}} \\
&= \frac{\prod_{j=1}^k \mathbf{E}\left[\mathrm{e}^{\lambda\cdot\mathbf{x}^j_i}\right]}{\mathrm{e}^{(1+\epsilon)\cdot\lambda k}} \\
&= \frac{\prod_{j=1}^k \mathrm{e}^{\lambda^2/2+\lambda}}{\mathrm{e}^{(1+\epsilon)\cdot\lambda k}} \\
&= \mathrm{e}^{k\lambda^2/2+k\lambda-(1+\epsilon)\lambda k} \\
&\leq \mathrm{e}^{-k\epsilon^2/2},
\end{aligned}$$

where the third equality follows by the moment generating function of normal distributions, and the last inequality holds by setting $\lambda = \epsilon$. Hence, in order to have $\mathrm{e}^{-k\epsilon^2/2} = O(1/d)$, we only need to ensure that $k = O(\log d/\epsilon^2)$.

For the second part, again assume we sample k points $\mathbf{x}^1, \ldots, \mathbf{x}^k \in \mathbb{R}^d$ such that $\mathbf{m_s} = \frac{1}{k}\sum_{j=1}^k \mathbf{x}^j$. Also note $\|\boldsymbol{\mu} - \mathbf{m_s}\|_2 \leq \epsilon$ if and only if $\|\boldsymbol{\mu} - \mathbf{m_s}\|_2^2 \leq \epsilon^2$. Let $z_i = \mu_i - (\mathbf{m_s})_i$ and also denote

$$Z := \|\mathbf{m_s} - \boldsymbol{\mu}\|^2 = \sum_{i=1}^d z_i^2.$$

We have that

$$Z = \sum_{i=1}^{d} \left(\frac{1}{k} \sum_{j=1}^{k} \mathbf{x}_i^j - \mu_i \right)^2 = \sum_{i=1}^{d} \left(\frac{1}{k} \left(\sum_{j=1}^{k} \mathbf{x}_i^j - k\mu_i \right) \right)^2 .$$

Now we know that

$$\mathbf{x}_i^j \sim \mathcal{N}(\mu_i, 1) \Rightarrow \sum_{j=1}^{k} \mathbf{x}_i^j \sim \mathcal{N}(k\mu_i, k) \Rightarrow \sum_{j=1}^{k} \mathbf{x}_i^j - k\mu_i \sim \mathcal{N}(0, k) \Rightarrow \frac{1}{k} \left(\sum_{j=1}^{k} \mathbf{x}_i^j - k\mu_i \right) \sim \mathcal{N}(0, 1/k).$$

Therefore $Z = \sum_{i=1}^{d} y_i^2$, where every $y_i \sim \mathcal{N}(0, 1/k)$. We can rearrange the previous equation to get

$$Z = \sum_{i=1}^{d} \frac{(\sqrt{k} y_i)^2}{k} \sim \frac{\chi^2(d)}{k}$$

To obtain our desired bound we apply Markov's inequality using the fact that $\mathbf{E}[Z] = \frac{d}{k}$ as follows:

$$\mathbf{P}[\|\boldsymbol{\mu} - \mathbf{m}_s\|_2 \geq \epsilon] = \mathbf{P}[Z \geq \epsilon^2] \leq \frac{\mathbf{E}[Z]}{\epsilon^2} = \frac{d}{k\epsilon^2}.$$

In other words, if we sample $k = O(d/\epsilon^2)$ points in $\mathbb{R}^d$ with very good probability the true mean $\boldsymbol{\mu}$ and the sampled mean $\mathbf{m_s}$ are very close in the ℓ_2-norm.

Problem 2: 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 $x = (1, 0, 0, \ldots, 0)$, 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 $x_1, \ldots, x_k \in \{0, 1\}$, 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.