15-859 Assignment #1

Albert Gao / sixiangg / section 1A

09/22/2022

Task 1

---

1. *Proof.* Let $z_i = x_i y_i$ then observe that
$$\begin{aligned}\langle x^{\otimes p}, y^{\otimes p}\rangle &= \sum_{i_1,i_2,\cdots i_p}\prod_{l\in[p]} x_{i_l}y_{i_l}\\ &= \sum_{i_1,i_2,\cdots,i_p}\prod_{l\in[p]} z_{i_l}\\ &= (z_1+\cdots+z_d)^p\\ &= \langle x,y\rangle^p.\end{aligned}$$
□

2. *Proof.* Let A_i denote ith row of A. Then for each $i\in[n]$, we have $M_i = A_i^{\otimes p/2}$. Then indeed M is a $n\times d^{p/2}$ matrix. It is also clear that
$$\begin{aligned}||Mx^{\otimes p/2}||_2^2 &= \sum_{i\in[n]}\langle M_i, x^{\otimes p/2}\rangle^2\\ &= \sum_{i\in[n]}\langle A_i, x\rangle^{p/2\times 2}\\ &= ||Ax||_p^p.\end{aligned}$$
To construct M, each entry is a $p/2$ product, so it takes $O(nd^{p/2}(p/2)) = n\cdot\text{poly}(d)$ time. Also, by discussion in class / problem statement, we may let S be a random $s\times n$ Gaussian matrix, where $s = O(d^{p/2}/\epsilon^2)$, such that with probability at least 9/10,
$$\begin{aligned}\forall x\in\mathbb{R}^{d^(p/2)}, ||SMx||_2^2 &= (1\pm\epsilon)||Mx||_2^2\\ \Rightarrow \forall x\in\mathbb{R}^d, ||SMx^{\otimes p/2}||_2^2 &= (1\pm\epsilon)||Mx^{\otimes p/2}||_2^2\\ &= (1\pm\epsilon)||Ax||_p^p.\end{aligned}$$
Finally, to compute SM we need only to perform $O(d^{p/2}/\epsilon^2)d^{p/2}$ dot products of vectors of length n, which takes in total $n\cdot\text{poly}(d)$ time. □

3. *Proof.* In the special case when A is Vandermonde, we no longer need all $d^{p/2}$ columns of M because of repetitive entries. Let S_k denote indices in $A_i^{\otimes p}$ for which the entries equal y_i^k. Then let M' have rows $(1, y_i, \cdots, y_i^{p(d-1)})$. Let y be a $p(d-1)+1$ length vector such that $y_k = \sum_{l\in S_k} x_l^{\otimes p/2}$. Then using same theorem in class / what the problem statement mentions, we can have matrix S appropriatedly scaled Gaussian with $O(p(d-1)/\epsilon^2) = O(dp/\epsilon^2)$ rows such that with at least 9/10 probability,
$$\begin{aligned}||SMx^{\otimes p/2}||_2^2 = ||SM'y||_2^2 &= (1\pm\epsilon)||M'y||_2^2\\ &= (1\pm\epsilon)||Mx^{\otimes p/2}||_2^2\\ &= (1\pm\epsilon)||Ax||_p^p.\end{aligned}$$
□

Task 2

Proof. First observe that

$$\begin{aligned}
||A-\widehat{A}||_2^2 &\le ||A-\widehat{A}||_F^2 \\
&\le \sum_{i,j}(A_{i,j}-\widehat{A}_{i,j})^2 \\
&\le \sum_{i,j}(\epsilon/2n)^2 \\
&= \epsilon^2/4 \\
\Rightarrow ||A-\widehat{A}||_2 &\le \epsilon/2.
\end{aligned}$$

Now, let $Y_t = \frac{\widehat{A}_{i_t,j_t}}{p_{i_t,j_t}} e_{i_t} e_{j_t}^T, X_t = \widehat{A} - Y_t$ so $\widehat{A} - \widetilde{A} = \frac{1}{s}\sum_{t=1}^s X_t$. We check that

$$\mathbb{E}[(X_t)_{i,j}] = \widehat{A}_{i,j} - p_{i,j}\frac{\widehat{A}_{i,j}}{p_{i,j}} = 0.$$

This also means $\mathbb{E}[Y_t] = \widehat{A}$. And with probability 1,

$$\begin{aligned}
||X_t||_2^2 &\le ||X_t||_F^2 \\
&\le ||\widehat{A}||_F^2 + \max_{i,j}\left(\widehat{A}_{i,j} - \frac{\widehat{A}_{i,j}}{p_{i,j}}\right)^2 - \widehat{A}_{i,j}^2 \\
&\le ||\widehat{A}||_F^2 + \max_{i,j}\widehat{A}_{i,j}^2(1/p_{i,j}-1) \\
&\le ||\widehat{A}||_F^2 + ||\widehat{A}||_F^2 = 2||\widehat{A}||_F^2 \\
\Rightarrow ||X_t||_2 &\le 2||\widehat{A}||_F.
\end{aligned}$$

So matrix Chernoff bound conditions are satisfied. We now bound the variance as follows.

$$\begin{aligned}
\left\|\sum_t \mathbb{E}[X_t^T X_t]\right\| &= \left\|s\mathbb{E}[X_t^T X_t]\right\| \\
&= s\|\widehat{A}^T\widehat{A} - \mathbb{E}[\widehat{A}^T Y_t] - \mathbb{E}[Y_t^T\widehat{A}] + \mathbb{E}[Y^T Y]\| \\
&\le s\left(\|\widehat{A}^T\widehat{A}\| + \|\mathbb{E}[\widehat{A}^T Y]\| + \|\mathbb{E}[Y_t^T\widehat{A}]\| + \|\mathbb{E}[Y^T Y]\|\right) \\
&= s\left(\|\widehat{A}^T\widehat{A}\| + \|\widehat{A}^T\widehat{A}\| + \|\widehat{A}^T\widehat{A}\| + \|B\|\right), B_{i,j} = \widehat{A}_{j,i}^2/p_{j,i} = ||\widehat{A}||_F^2 \\
&\le s\left(3\|\widehat{A}^T\widehat{A}\|_F + \|B\|_F\right) \\
&\le s\left(3\|\widehat{A}^T\|_F\|\widehat{A}\|_F + \sqrt{n^2\|\widehat{A}\|_F^4}\right) \\
&= s\left(3\|\widehat{A}\|_F^2 + n\|\widehat{A}\|_F^2\right).
\end{aligned}$$

Similarly we have $\|\sum_t \mathbb{E}[X_t X_t^T]\| \le s\left(3\|\widehat{A}\|_F^2 + n\|\widehat{A}\|_F^2\right)$, and we finally can use the matrix Chernoff bound from wikipedia to get

$$\begin{aligned}
\Pr[\|\widetilde{A} - \widehat{A}\|_2 \ge \epsilon/2] &= \Pr[\frac{1}{s}\|\sum X_t\|_2 \ge \epsilon/2] \\
&= \Pr[\|\sum X_t\|_2 \ge s\epsilon/2] \\
&\le (2n)\exp\left(-\frac{s^2\epsilon^2/8}{\sigma^2 + 2\|A\|_F s\epsilon/6}\right) \\
&\le (2n)\exp\left(-\frac{s^2\epsilon^2/8}{s(3+n)\|\widehat{A}\|_F^2 + \|\widehat{A}\|_F s\epsilon/3}\right) \\
&\le (2n)\exp\left(-\frac{\Omega(\|\widehat{A}\|_F^2 n\epsilon^{-2}\ln n)\epsilon^2/8}{(3+n)\|\widehat{A}\|_F^2 + \|\widehat{A}\|_F\epsilon/3}\right) \\
&= (2n)\exp\left(-\frac{\Omega(\|\widehat{A}\|_F n\ln n)}{(3+n)\|\widehat{A}\|_F + \epsilon/3}\right) \\
&\le (2n)\exp\left(-\frac{\Omega(\|\widehat{A}\|_F n\ln n)}{(3+n)\|\widehat{A}\|_F + 2n\|\widehat{A}\|_F/3}\right) \\
&\le (2n)\exp\left(-\Omega(\ln n)\right) \\
&= (2n)\exp\left(-\Omega(\ln n^2)\right) \\
&= (2n)O(1/n^2) \\
&= O(1/n).
\end{aligned}$$

It is clear that we can set the constants correctly such that the probability is at most $1/n$. Hence we high probability (at least $1 - 1/n$) we get

$$\|A - \widetilde{A}\|_2 \le \|A - \widehat{A}\|_2 + \|\widehat{A} - \widetilde{A}\|_2 \le \epsilon/2 + \epsilon/2 = \epsilon.$$

We're done. $\square$

Task 3

Proof. As hint says, we should first prove for vectors.

<u>Claim:</u> $\forall i, j, |(AB)_{i,j} - (\overline{AB})_{i,j}| \le \epsilon \|A_{i*}\|_1 \|B_{*j}\|_1$.

Proof. Note that for any k,

$$|\overline{A}_{ik}\overline{B}_{kj} - A_{ik}B_{kj}| = \begin{cases} 0, & |A_{ik}| > (\epsilon/2)\|A_{i*}\|_1, |B_{kj}| > (\epsilon/2)\|B_{*j}\|_1 \\ |A_{ik}B_{kj}|, & \text{o.w.} \end{cases}.$$

Then we have

$$\begin{aligned} |(AB)_{i,j} - (\overline{AB})_{i,j}| &\le \sum_k |\overline{A}_{ik}\overline{B}_{kj} - A_{ik}B_{kj}| \\ &\le \sum_k (\epsilon/2) \max\left\{ |A_{ik}| \sum_t |B_{tj}|, \sum_t |A_{it}||B_{kj}| \right\} \\ &\le (\epsilon/2) \sum_k \sum_t 2|A_{ik}||B_{tj}| \\ &= \epsilon \|A_{i*}\|_1 \|B_{*j}\|_1. \end{aligned}$$

Note that the last inequality follows from the observation that each $|A_{ik}||B_{tj}||$ (for each k, t) is summed at most twice, once when the outer index is at k and once when the outer index is at t. □

Now, it is clear that

$$\begin{aligned} \|AB - \overline{A} \cdot \overline{B}\|_1 &= \sum_{i,j} |(AB)_{ij} - (\overline{A} \cdot \overline{B})_{ij}| \\ &\le \epsilon \sum_{i,j} \|A_{i*}\|_1 \|B_{*j}\|_1 \\ &= \epsilon \sum_i \|A_{i*}\|_1 \sum_j \|B_{*j}\|_1 \\ &= \epsilon \|A\|_1 \|B\|_1. \end{aligned}$$

□

Task 4

Proof. Our algorithm is as follows.

> Use $36 \log m$ independent count sketch matrices, each containing $s = 3600d^2$ rows.
> For each such matrix S_k :
>   Let $C = S_k A, D = S_k B$.
>   Let $L \in \mathbb{R}^{O(\log m)\times s}$ drawn from i.i.d. Gaussians satisfying distributional JL lemma.
>   Let $M = LCC^{-}D - LD$ via matrix multiplications left-to-right.
>   Let $v_k \in \mathbb{R}^m$ with $(v_k)_i = \|M_{*i}\|_2$.
> Let $v \in \mathbb{R}^m$ whose entries are medians of $\{(v_k)_i\}_{k\in[s]}$.
> Return $\min_i v_i$

Claim: For all columns j, with 5/6 probability, $\|M_{*j}\|_2$ approximates optimal $\|Cx - D_{*j}\|_2$ within 1 ± 0.1.

Proof. Let $\epsilon = 0.1$. Let $\delta = \frac{1}{6m}$. Since L matrix has $O(\epsilon^{-2}\log(1/\delta)) = O(\log m)$ rows, it satisfies JL property (e.g., from wikipedia). Let $w_j = (CC^{-}D)_{*j} - D_{*j}$, the solution to $\min_x \|Cx - D_{*j}\|_2$. By JL property, we have

$$\begin{aligned}
&\forall j \in [m], \Pr[\|Lw_j\|_2 - \|w_j\|_2 > \|w_j\|_2\epsilon] < \delta \\
\Rightarrow &\Pr\left[\exists j \in [m], \|Lw_j\|_2 - \|w_j\|_2 > \|w_j\|_2\epsilon\right] < m\delta \\
\Rightarrow &\Pr\left[\forall j \in [m], \|Lw_j\|_2 - \|w_j\|_2 > \|w_j\|_2\epsilon\right] > 1 - m\delta \\
\Rightarrow &\Pr\left[\forall j \in [m], \|M_{*j}\|_2 \in (1 \pm \epsilon)\|w_j\|_2\right] > 5/6.
\end{aligned}$$

□

Claim: For a fixed column j, with 2/3 probability, $\|M_{*j}\|_2$ approximates optimal $\|Ax - B_{*j}\|_2$ within a factor of 2.

Proof. Let $\epsilon = 0.1$, $\delta = 1/6$. Since $s = 3600d^2 = 6d^2/(\delta\epsilon^2)$, our count sketch matrix satisfies subspace embedding. Let $r_j^* = \min_x \|Ax - B_{*j}\|_2$. Then

$$\begin{aligned}
\Pr[\|M_{*j}\|_2 \notin (1/2, 2)r_j^*] &\le \Pr[\|M_{*j}\|_2 \notin (1 \pm \epsilon)\|w_j\|_2] + \Pr[\|w_j\|_2 \notin (1 \pm \epsilon)r_j^*] \\
&\le 1/6 + 1/6 \\
&= 1/3.
\end{aligned}$$

□

Claim: We approximate least cost regression within a factor of 2, with probability at least 9/10.

Proof. In order for the median of $t = 36 \log m$ values to fail (for some fixed j), we need at least half of them to fail (where failing means landing outside the factor of 2 range). Each has failure probability at most $\frac{1}{3}$, so if we let $X_{i\in[t]} = 1$ iff failure, by Hoeffding we then have

$$\begin{aligned}
\Pr[\sum X_i \geq t/2] &= \Pr[\sum X_i - \frac{1}{3}t \geq \frac{1}{6}t] \\
&\leq \Pr[\sum X_i - \mathbb{E}[\sum X_i] \geq \frac{1}{6}t] \\
&\leq \exp\left(-\frac{t^2/18}{t}\right) \\
&= e^{-t/18} \\
&= m^{-2}
\end{aligned}$$

By union bound we know that probability that at least one failure occurs out of m columns is at most m^{-1}, and so all successful approximation has probability at least $1 - 1/m$. We may assume $m \geq 10$ here since otherwise we can just solve all regression problems and take minimum naively. It is also easy to see that if for all columns, our approximation is at least within a factor of 2, the minimum is also within a factor of 2. We're done. □

Finally, we show our algorithm runs quickly. Note that for each countsketch matrix, computing C, D uses $nnz(A) + nnz(B)$ time and drawing Gaussians uses $O(d^2 \log m)$ time. Since $nnz(C) \leq nnz(A), nnz(D) \leq nnz(B)$ we have LC taking at most $O(nnz(A) \log m)$ time, C^- taking $O(d^2 d^2) = O(d^4)$ time, $(LC)C^-$ taking $O(dd^2 \log m) = O(d^3 \log m)$ time, $(LCC^-)D$ and computing LD taking $O(\log(m) nnz(B))$ time, and computing M taking $O(m \log m)$ time. Figuring out column norms of M takes $O(m \log m)$ time. So for each S_k our work is

$$\begin{aligned}
&O(nnz(A) + nnz(B) + d^2 \log m + nnz(A) \log m + d^4 + d^3 \log m + nnz(B) \log m + 2m \log m) \\
=&O(\log m(nnz(A) + nnz(B) + d^4))
\end{aligned}$$

assuming that $m \leq nnz(B)$ since otherwise we can just cut the zero columns of B as a preprocessing step. Taking elem-wise median of $\log m$ vectors of length m takes time $m\mathsf{polylog}(m)$ so in total our algo runs in time

$$\begin{aligned}
&O(\log m) O(\log m(nnz(A) + nnz(B) + d^4)) + O(m\mathsf{polylog}(m)) \\
=&O(\mathsf{polylog}(m)(nnz(A) + nnz(B) + d^4))
\end{aligned}$$

as desired. [Oh man that was crazy.] □