School of Computer Science and Technology
University of Science and Technology of China

Exercise Sheet 2 for
Algorithms for Big Data
2023 Spring
Solution

Note: In the following, for a vector $\boldsymbol{x}$, $\|\boldsymbol{x}\| = \|\boldsymbol{x}\|_2$.

Exercise 1 (15 points, graded by Jing Cao)
Let $\sum_{i=1}^{r} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^\top$ be the SVD of A, where $A \in \mathbb{R}^{n\times d}$. Show that $\left\|\boldsymbol{u}_1^\top A\right\| = \sigma_1$ and $\left\|\boldsymbol{u}_1^\top A\right\| = \max_{\|\boldsymbol{u}\|=1} \left\|\boldsymbol{u}^\top A\right\|$, where $\|\boldsymbol{x}\| = \sqrt{\sum_{i=1}^{d} x_i^2}$ for a vector $\boldsymbol{x} \in \mathbb{R}^d$.

Proof. We first calculate $\left\|\boldsymbol{u}_1^\top A\right\|$. Because the left-singular vectors are pairwise orthogonal, we have $\boldsymbol{u}_1^\top \boldsymbol{u}_1 = 1$ and $\boldsymbol{u}_1^\top \boldsymbol{u}_i = 0$ for all $2 \le i \le r$. Thus

$$\boldsymbol{u}_1^\top A = \boldsymbol{u}_1^\top \sum_{i=1}^{r} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^\top = \sigma_1 (\boldsymbol{u}_1^\top \boldsymbol{u}_1) \boldsymbol{v}_1 + \sum_{i=2}^{r} \sigma_i (\boldsymbol{u}_1^\top \boldsymbol{u}_i) \boldsymbol{v}_i^\top = \sigma_1 \boldsymbol{v}_1^\top.$$

From the definition of right singular vectors we can see that $\|\boldsymbol{v}_1\| = 1$. Thus

$$\left\|\boldsymbol{u}_1^\top A\right\| = \left\|\sigma_1 \boldsymbol{v}_1^\top\right\| = \sigma_1.$$

Next we will show that $\sigma_1 = \max_{\|\boldsymbol{u}\|=1} \left\|\boldsymbol{u}^\top A\right\|$. First we extend $\{\boldsymbol{u}_i\}$ to an orthonormal basis of $\mathbb{R}^n$, i.e. we recursively choose $\boldsymbol{u}_i$ to be some unit vector that is perpendicular to $U_{i-1} = \operatorname{span}\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_{i-1}\}$ for all $i = r+1, r+2, \ldots, n$. For any vector $\boldsymbol{u} \in \mathbb{R}^n$, we expand $\boldsymbol{u}$ on the basis $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_r\}$, denote it as $\boldsymbol{u} = \sum_{i=1}^{n} x_i \boldsymbol{u}_i$. Since $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_r\}$ is an orthonormal basis, we have $\sum_{i=1}^{n} x_i^2 = 1$ and

$$\boldsymbol{u}^\top A = \left(\sum_{i=1}^{n} x_i \boldsymbol{u}_i^\top\right)\left(\sum_{i=1}^{r} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^\top\right) = \sum_{i=1}^{r} x_i \sigma_i \boldsymbol{v}_i^\top,$$

$$\left\|\boldsymbol{u}^\top A\right\|^2 = \sum_{i=1}^{r} x_i^2 \sigma_i^2 = \sigma_1^2 \sum_{i=1}^{r} x_i^2 - \sum_{i=1}^{r} (\sigma_1^2 - \sigma_i^2) x_i^2 \le \sigma_1^2 \sum_{i=1}^{r} x_i^2 = \sigma_1^2.$$

Also notice that when $x_1 = 1$ and $x_2 = x_3 = \cdots = x_n = 0$, we have $\left\|\boldsymbol{u}^\top A\right\| = \sigma_1$. Thus $\left\|\boldsymbol{u}_1^\top A\right\| = \sigma_1 = \max_{\|\boldsymbol{u}\|=1} \left\|\boldsymbol{u}^\top A\right\|$. $\square$

Exercise 2 (25 points, graded by Zelin Li)
Let A be an $n \times d$ matrix with SVD such that $A = \sum_{i=1}^{r} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^\top$. Let $\boldsymbol{x} \in \mathbb{R}^d$ be a vector such that $\|\boldsymbol{x}\|_2 = 1$ and $\left|\boldsymbol{x}^\top \boldsymbol{v}_1\right| \ge \delta$ for some $\delta > 0$. Suppose that $\sigma_2 < \frac{1}{2}\sigma_1$. Let $\boldsymbol{w}$ be the vector after $k = \log(1/\varepsilon\delta)$ iterations of the power method, namely,

$$\boldsymbol{w} = \frac{\left(A^\top A\right)^k \boldsymbol{x}}{\left\|\left(A^\top A\right)^k \boldsymbol{x}\right\|_2}.$$

Prove that the length of the projection of $\boldsymbol{w}$ onto the line defined by the first singular vector $\boldsymbol{v}_1$ is at least $1-\varepsilon$, i.e., $\left|\boldsymbol{w}^\top \boldsymbol{v}_1\right| \ge 1-\varepsilon$.

Proof. We add $\boldsymbol{v}_{r+1}, \boldsymbol{v}_{r+2}, \ldots, \boldsymbol{v}_d$ so that $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_d\}$ forms an orthonormal basis of $\mathbb{R}^d$. Write $\boldsymbol{x}$ as the linear combination of $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_d\}$ and we have

$$A^\top A\boldsymbol{x} = \left(\sum_{i=1}^{r} \sigma_i \boldsymbol{v}_i \boldsymbol{u}_i^\top\right)\left(\sum_{i=1}^{r} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^\top\right)\left(\sum_{i=1}^{d} a_i \boldsymbol{v}_i\right) = \sum_{i=1}^{r} a_i \sigma_i^2 \boldsymbol{v}_i, \tag{1}$$

where $a_i = \langle \boldsymbol{x}, \boldsymbol{v}_i \rangle$. According to (1), we have

$$\left(A^\top A\right)^k \boldsymbol{x} = \sum_{i=1}^{r} a_i \sigma_i^{2k} \boldsymbol{v}_i,$$

$$\left(\left(A^\top A\right)^k \boldsymbol{x}\right)^\top \boldsymbol{v}_1 = \left(\sum_{i=1}^{r} a_i \sigma_i^{2k} \boldsymbol{v}_i\right)^\top \boldsymbol{v}_1 = a_1 \sigma_1^{2k}.$$

To prove $\left|\boldsymbol{w}^\top \boldsymbol{v}_1\right| \geq 1 - \varepsilon$, i.e., $\left|\left(\left(A^\top A\right)^k \boldsymbol{x}\right)^\top \boldsymbol{v}_1\right| \geq (1-\varepsilon)\left\|\left(A^\top A\right)^k \boldsymbol{x}\right\|_2$, we need to bound $\left\|\left(A^\top A\right)^k \boldsymbol{x}\right\|_2$. By calculus, we have

$$\begin{aligned}
\left\|\left(A^\top A\right)^k \boldsymbol{x}\right\|_2 &= \sqrt{\sum_{i=1}^{r} a_i^2 \sigma_i^{4k}} \\
&< \sqrt{a_1^2 \sigma_1^{4k} + \left(\frac{1}{2}\right)^{4k} \sigma_1^{4k} \sum_{i=2}^{r} a_i^2} \\
&\leq a_1 \sigma_1^{2k} \sqrt{1 + \left(\frac{1}{2}\right)^{4k} \left(\frac{1}{\delta^2} - 1\right)}.
\end{aligned}$$

The last inequality comes from $\sum_{i=2}^{r} a_i^2 \leq 1 - a_1^2 \leq a_1^2\left(\frac{1}{a_1^2} - 1\right) \leq a_1^2\left(\frac{1}{\delta^2} - 1\right)$. Take $k = \log(1/\varepsilon\delta)$, we have

$$\begin{aligned}
(1-\varepsilon)\left\|\left(A^\top A\right)^k \boldsymbol{x}\right\|_2 &< a_1 \sigma_1^{2k} (1-\varepsilon)\sqrt{1 + \varepsilon^4 \delta^2} \\
&< a_1 \sigma_1^{2k} (1-\varepsilon)\left(1 + \frac{\varepsilon^4 \delta^2}{2}\right) \\
&< a_1 \sigma_1^{2k} (1-\varepsilon)(1+\varepsilon) \\
&< a_1 \sigma_1^{2k} = \left|\left(\left(A^\top A\right)^k \boldsymbol{x}\right)^\top \boldsymbol{v}_1\right|.
\end{aligned}$$

Here we use the fact that $\sqrt{1+x} < 1 + \frac{x}{2}$ when $x > 0$, and $0 < \varepsilon, \delta < 1$. □

Exercise 3 (20 points, graded by Jing Cao)
Let $k < d$. Let $U \in \mathbb{R}^{d \times k}$ be a random matrix such that its (i,j)-th entry is denoted as u_{ij}, where $\{u_{ij}\}$ are independent random variables such that

$$u_{ij} = \begin{cases} 1 & \text{with probability } \frac{1}{2}, \\ -1 & \text{with probability } \frac{1}{2}. \end{cases}$$

Now we use matrix U as a random projection matrix. That is, for a (row) vector $\boldsymbol{a} \in \mathbb{R}^d$, we map it to

$$f(\boldsymbol{a}) = \frac{1}{\sqrt{k}}\boldsymbol{a}U.$$

For each j such that $1 \leq j \leq k$, define $b_j = [f(\boldsymbol{a})]_j$, i.e., b_j is the j-th entry of $f(\boldsymbol{a})$.

(1) (8 points) What is the expectation $\mathrm{E}[b_j]$?

(2) (8 points) What is $\mathrm{E}[b_j^2]$?

(3) (4 points) What is $\mathrm{E}[\|f(\boldsymbol{a})\|^2]$?

Solution.

(1) We first calculate that $\mathrm{E}[u_{ij}] = 1 \cdot \Pr[u_{ij} = 1] + (-1) \cdot \Pr[u_{ij} = -1] = 1 \times \frac{1}{2} + (-1) \times \frac{1}{2} = 0$.
Since $b_j = [f(\boldsymbol{a})]_j = \frac{1}{\sqrt{k}} \sum_{i=1}^d a_i u_{ij}$, and by linearity of expectation, we have

$$\mathrm{E}[b_j] = \mathrm{E}\left[\frac{1}{\sqrt{k}} \sum_{i=1}^d a_i u_{ij}\right] = \frac{1}{\sqrt{k}} \sum_{i=1}^d a_i \mathrm{E}[u_{ij}] = 0.$$

(2) Since $\mathrm{Var}[b_j] = \mathrm{E}[b_j^2] - (\mathrm{E}[b_j])^2 = \mathrm{E}[b_j^2] - 0 = \mathrm{E}[b_j^2]$, it suffices to calculate $\mathrm{Var}[b_j]$.
We first calculte that $\mathrm{Var}[u_{ij}] = \mathrm{E}[u_{ij}^2] - (\mathrm{E}[u_{ij}])^2 = \mathrm{E}[u_{ij}^2] = 1$. Since $\{u_{ij}\}$ are independent random variables, we have

$$\mathrm{E}[b_j^2] = \mathrm{Var}[b_j] = \mathrm{Var}\left[\frac{1}{\sqrt{k}} \sum_{i=1}^d a_i u_{ij}\right] = \frac{1}{k} \sum_{i=1}^d a_i^2 \mathrm{Var}[u_{ij}] = \frac{1}{k} \sum_{i=1}^d a_i^2 = \frac{\|\boldsymbol{a}\|^2}{k}.$$

(3) Since $\|f(\boldsymbol{a})\|^2 = \sum_{j=1}^k b_j^2$, and by linearity of expectation, we have

$$\mathrm{E}\left[\|f(\boldsymbol{a})\|^2\right] = \mathrm{E}\left[\sum_{j=1}^k b_j^2\right] = \sum_{j=1}^k \mathrm{E}[b_j^2] = k \cdot \frac{\|\boldsymbol{a}\|^2}{k} = \|\boldsymbol{a}\|^2.$$

Exercise 4 (20 points, graded by Yinhao Dong)
In the class, we have seen an algorithm, denoted by $\mathcal{A}$, for the (c, r)-ANN problem with success probability at least 0.6. That is, upon a queried vertex x such that there exists a point a^* in the set $\mathcal{P}$ with $d(x, a^*) \leq r$, the algorithm $\mathcal{A}$ outputs some $a \in \mathcal{P}$ with $d(x, a) \leq c \cdot r$ with probability at least 0.6.
Let $\delta \in (0, 1)$. Using the above $\mathcal{A}$ as a subroutine, give a new algorithm $\mathcal{B}$ with success probability at least $1 - \delta$. That is, for the above query vertex x, the algorithm $\mathcal{B}$ outputs some $a \in \mathcal{P}$ with $d(x, a) \leq c \cdot r$ with probability at least $1 - \delta$. Your algorithm should use as little query time as possible. Explain the correctness of your algorithm and state its query time, assuming the query time of $\mathcal{A}$ is $T_{\mathcal{A}}$.

Solution.
The algorithm $\mathcal{B}$ is given as follows (10 points):

(a) Independently initialize $k \triangleq \lceil \log_{0.4} \delta \rceil$ copies of $\mathcal{A}$.

(b) Upon a query, iteratively query every subroutine:

- If one of them outputs a point a with $d(x, a) \leq c \cdot r$, then output a.
- If all of them output a with $d(x, a) > c \cdot r$, then output FAIL.

Correctness (5 points): For every subroutine, since the probability that it succeeds in outputing some (c, r)-ANN is at least 0.6, the probability that it fails is 0.4. Since the subroutines are independent, the probability that all of them fail is at most 0.4^k. Since we run $k = \lceil \log_{0.4} \delta \rceil \geq \log_{0.4} \delta$ copies of $\mathcal{A}$, we have $0.4^k \leq 0.4^{\log_{0.4} \delta} = \delta$. Thus the probability that algorithm $\mathcal{B}$ fails is at most δ.

Query time (5 points): $k \cdot T_{\mathcal{A}} = \lceil \log_{0.4} \delta \rceil \cdot T_{\mathcal{A}}$.

Exercise 5 (20 points, graded by Xiaoyang Xu)
Let $\alpha \in (0, 1]$. Suppose we change the (basic) Morris algorithm to the following:

(a) Initialize $X \leftarrow 0$.

(b) For each update, increment X by 1 with probability $\frac{1}{(1+\alpha)^X}$.

(c) For a query, output $\tilde{n} = \frac{(1+\alpha)^X - 1}{\alpha}$.

Let X_n denote X in the above algorithm after n updates. Let $\tilde{n} = \frac{(1+\alpha)^{X_n} - 1}{\alpha}$.

(1) (10 points) Calculate $\mathrm{E}[\tilde{n}]$ and upper bound $\mathrm{Var}[\tilde{n}]$.

Solution.
We first calculate $\mathrm{E}[\tilde{n}]$. By definition, we have $\Pr[X_{n+1} = x + 1 \mid X_n = x, x \in \mathbb{N}] = \frac{1}{(1+\alpha)^x}$ and $\Pr[X_{n+1} = x \mid X_n = x, x \in \mathbb{N}] = 1 - \frac{1}{(1+\alpha)^x}$. Thus

$$
\begin{aligned}
&\mathrm{E}\left[(1+\alpha)^{X_{n+1}}\right] \\
=&\sum_{x\in\mathbb{N}} (1+\alpha)^x \Pr\left[X_{n+1} = x\right] \\
=&\sum_{x\in\mathbb{N}} (1+\alpha)^x \Pr\left[X_{n+1} = x \wedge (X_n = x \vee X_n = x-1)\right] \\
=&\sum_{x\in\mathbb{N}} (1+\alpha)^x (\Pr\left[X_{n+1} = x \wedge X_n = x\right] + \Pr[X_{n+1} = x \wedge X_n = x-1]) \\
=&\sum_{x\in\mathbb{N}} \left((1+\alpha)^x \Pr\left[X_{n+1} = x \wedge X_n = x\right] + (1+\alpha)^{x+1} \Pr[X_{n+1} = x+1 \wedge X_{n+1} = x]\right) \\
=&\sum_{x\in\mathbb{N}} \left((1+\alpha)^x \Pr\left[X_{n+1} = x \mid X_n = x\right] + (1+\alpha)^{x+1} \Pr\left[X_{n+1} = x+1 \mid X_n = x\right]\right) \Pr[X_n = x] \\
=&\sum_{x\in\mathbb{N}} \left((1+\alpha)^x \left(1 - \frac{1}{(1+\alpha)^x}\right) + (1+\alpha)^{x+1} \cdot \frac{1}{(1+\alpha)^x}\right) \Pr[X_n = x] \\
=&\sum_{x\in\mathbb{N}} ((1+\alpha)^x \Pr[X_n = x] + \alpha \Pr[X_n = x]) \\
=&\sum_{x\in\mathbb{N}} (1+\alpha)^x \Pr[X_n = x] + \alpha \sum_{x\in\mathbb{N}} \Pr[X_n = x] \\
=&\mathrm{E}\left[(1+\alpha)^{X_n}\right] + \alpha.
\end{aligned}
$$

Since $X_0 = 0$, i.e., $\mathrm{E}[(1+\alpha)^{X_0}] = 1$, it is easy to prove by induction that $\mathrm{E}[(1+\alpha)^{X_n}] = \alpha n + 1$. By definition we have $\tilde{n} = \frac{(1+\alpha)^{X_n} - 1}{\alpha}$. Hence $\mathrm{E}[\tilde{n}] = \frac{\alpha n + 1 - 1}{\alpha} = n$.

We then calculate $\mathrm{Var}[\tilde{n}]$. We use the same method as above to calculate $\mathrm{E}\left[(1+\alpha)^{2X_n}\right]$.

$$
\begin{aligned}
&\mathrm{E}\left[(1+\alpha)^{2X_{n+1}}\right]\\
=&\sum_{x\in\mathbb{N}}(1+\alpha)^{2x}\Pr\left[X_{n+1}=x\right]\\
=&\sum_{x\in\mathbb{N}}(1+\alpha)^{2x}\Pr\left[X_{n+1}=x\wedge(X_n=x\vee X_n=x-1)\right]\\
=&\sum_{x\in\mathbb{N}}(1+\alpha)^{2x}(\Pr\left[X_{n+1}=x\wedge X_n=x\right]+\Pr[X_{n+1}=x\wedge X_n=x-1])\\
=&\sum_{x\in\mathbb{N}}\left((1+\alpha)^{2x}\Pr\left[X_{n+1}=x\wedge X_n=x\right]+(1+\alpha)^{2x+2}\Pr[X_{n+1}=x+1\wedge X_{n+1}=x]\right)\\
=&\sum_{x\in\mathbb{N}}\left((1+\alpha)^{2x}\Pr\left[X_{n+1}=x\mid X_n=x\right]+(1+\alpha)^{2x+2}\Pr\left[X_{n+1}=x+1\mid X_n=x\right]\right)\Pr[X_n=x]\\
=&\sum_{x\in\mathbb{N}}\left((1+\alpha)^{2x}\left(1-\frac{1}{(1+\alpha)^x}\right)+(1+\alpha)^{2x+2}\cdot\frac{1}{(1+\alpha)^x}\right)\Pr[X_n=x]\\
=&\sum_{x\in\mathbb{N}}\left((1+\alpha)^{2x}\Pr[X_n=x]+(2\alpha+\alpha^2)(1+\alpha)^x\Pr[X_n=x]\right)\\
=&\sum_{x\in\mathbb{N}}(1+\alpha)^{2x}\Pr[X_n=x]+(2\alpha+\alpha^2)\sum_{x\in\mathbb{N}}(1+\alpha)^x\Pr[X_n=x]\\
=&\mathrm{E}\left[(1+\alpha)^{2X_n}\right]+(2\alpha+\alpha^2)\mathrm{E}\left[(1+\alpha)^{X_n}\right].
\end{aligned}
$$

Since $\tilde{n}^2=\frac{(1+\alpha)^{2X_n}-2(1+\alpha)^{X_n}+1}{\alpha^2}$, we can see that

$$
\begin{aligned}
\mathrm{E}[\widetilde{n+1}^2]&=\mathrm{E}\left[\frac{(1+\alpha)^{2X_{n+1}}-2(1+\alpha)^{X_{n+1}}+1}{\alpha^2}\right]\\
&=\frac{1}{\alpha^2}\mathrm{E}\left[(1+\alpha)^{2X_{n+1}}\right]-\frac{2}{\alpha^2}\mathrm{E}[(1+\alpha)^{X_{n+1}}]+\frac{1}{\alpha^2}\\
&=\frac{1}{\alpha^2}(\mathrm{E}[(1+\alpha)^{2X_n}]+(2\alpha+\alpha)^2\mathrm{E}[(1+\alpha)^{X_n}])-\frac{2}{\alpha^2}(\mathrm{E}[(1+\alpha)^{X_n}]+\alpha)+\frac{1}{\alpha^2}\\
&=\left(\frac{1}{\alpha^2}\mathrm{E}\left[(1+\alpha)^{2X_n}\right]-\frac{2}{\alpha^2}\mathrm{E}[(1+\alpha)^{X_n}]+\frac{1}{\alpha^2}\right)+\frac{2+\alpha}{\alpha}\mathrm{E}[(1+\alpha)^{X_n}]-\frac{2}{\alpha}\\
&=\mathrm{E}[\tilde{n}]+\frac{2+\alpha}{\alpha}(\alpha n+1)-\frac{2}{\alpha}\\
&=\mathrm{E}[\tilde{n}]+(\alpha+2)n+1.
\end{aligned}
$$

We can prove by induction that $\mathrm{E}[\tilde{n}^2]=\frac{2+\alpha}{2}n^2-\frac{\alpha}{2}n$. Hence $\mathrm{Var}[\tilde{n}]=\mathrm{E}[\tilde{n}^2]-(\mathrm{E}[\tilde{n}])^2=\frac{\alpha}{2}(n^2-n)<\frac{\alpha}{2}n^2$.

For each of $\mathrm{E}[\tilde{n}]$ and $\mathrm{Var}[\tilde{n}]$, you get 5 points if you calculate the correct answer and can prove the correctness of your calculation. Otherwise, if you can find the relationship between $\widetilde{n+1}$ and $\tilde{n}$, you get 3 points.

(2) (10 points) Let $\varepsilon,\delta\in(0,1)$. Based upon the above algorithm (you may select whatever α as you wish), give a new algorithm such that with probability at least $1-\delta$, it outputs an estimator $\tilde{n}$ such that $|\tilde{n}-n|\le\varepsilon n$. Explain the correctness and the (*worst-case*) space complexity (i.e., the number of bits) of your algorithm. It suffices to give an algorithm such that with probability at least $1-\delta'$, its *worst-case* space complexity is a polynomial function of $\frac{1}{\delta}$, $\frac{1}{\delta'}$, $\frac{1}{\varepsilon}$ and $\log\log n$, i.e., $\mathrm{poly}(\frac{1}{\delta},\frac{1}{\delta'},\frac{1}{\varepsilon},\log\log n)$.

Solution.
Algorithm description (5 points):

(a) Independently run $k \triangleq \left\lceil \frac{\alpha}{2\delta\varepsilon^2} \right\rceil$ copies of the modified Morris algorithm. Update each of the subroutines when requested to update.

(b) Upon a query, query each of these subroutines and let $\tilde{n}_1, \tilde{n}_2, \ldots, \tilde{n}_k$ denote their outputs. Output their arithmetic mean, i.e., $\tilde{n} = \frac{1}{k}\sum_{i=1}^{k} \tilde{n}_i$.

Correctness (2 points):

By (1), after n updates, we have $\mathrm{E}[\tilde{n}_1] = \cdots = \mathrm{E}[\tilde{n}_k] = n$ and $\mathrm{Var}(\tilde{n}_1) = \cdots = \mathrm{Var}(\tilde{n}_k) < \frac{\alpha}{2}n^2$. Since $\tilde{n} = \frac{1}{k}\sum_{i=1}^{k} \tilde{n}_i$ and $\tilde{n}_i$'s are independent variables, we have $\mathrm{E}[\tilde{n}] = n$ and $\mathrm{Var}[\tilde{n}] = \frac{1}{k^2}\sum_{i=1}^{k} \mathrm{Var}[\tilde{n}_i] < \frac{\alpha}{2k}n^2$.

By Chebyshev's inequality, since $k \geq \frac{\alpha}{2\delta\varepsilon}$, we have

$$\Pr\left[|\tilde{n} - n| > \varepsilon n\right] \leq \frac{\mathrm{Var}[\tilde{n}]}{\varepsilon^2 n^2} < \frac{\alpha n^2}{2k\varepsilon^2 n^2} \leq \frac{\alpha n^2}{2\frac{\alpha}{2\delta\varepsilon^2}\varepsilon^2 n^2} = \delta.$$

This also indicates that with probability at least $1-\delta$, we have $|\tilde{n} - n| \leq \varepsilon n$.

Space complexity (3 points):

We first analyze the space complexity of every subroutine.

Since each subroutine is only increased n times, i.e., each counter X_i only have n chances to increase, the probability that $X_i > \left\lceil \log_{1+\alpha} \frac{nk}{\delta'} \right\rceil$ is at most $\frac{\delta'}{k}$, as the probability that X_i increases at each update is at most $\frac{1}{(1+\alpha)^{\left\lceil \frac{nk}{\delta'} \right\rceil}} \leq \frac{\delta'}{nk}$.

By union bound, we can see that the probability that X_i increases at least once after it reaches $\left\lceil \log_{1+\alpha} \frac{nk}{\delta'} \right\rceil$ is at most $\frac{\delta'}{k}$, which indicates that with probability greater than $1 - \frac{\delta'}{k}$ we have $X_i \leq \left\lceil \log_{1+\alpha} \frac{nk}{\delta'} \right\rceil$. This means that the space complexity of each subroutine is $O(\log_{1+\alpha} \frac{nk}{\delta}) = O(\log_{1+\alpha} n + \log_{1+\alpha} k + \log_{1+\alpha} \frac{1}{\delta'})$ with probability at least $1 - \frac{\delta'}{k}$.

As we run k same subroutines and the probability that each subroutine breaks the space complexity upper bound $O(\log_{1+\alpha} n + \log_{1+\alpha} k + \log_{1+\alpha} \frac{1}{\delta'})$ is at most $\frac{\delta'}{k}$, by unoin bound again, we can see that with probability at most δ', some subroutine breaks the bound above, and thus with probability greater than $1-\delta'$, no subroutine breaks this bound. This indicates that with probability greater than $1-\delta'$, the space complexity of the entire algorithm is $kO\left(\log(\log_{1+\alpha} n + \log_{1+\alpha} k + \log_{1+\alpha} \frac{1}{\delta'})\right)$.

Since $k = O(\frac{\alpha}{2\delta\varepsilon^2})$ and α is only some constant, we can give a upper bound of the space complexity of the entire algorithm

$$O\left(\frac{1}{\delta}\cdot\frac{1}{\varepsilon^2}\log\left(\log n + \log\frac{1}{\delta\varepsilon^2} + \log\frac{1}{\delta'}\right)\right) = O\left(\frac{1}{\delta}\cdot\frac{1}{\varepsilon^2}\left(\log\log n + \log\log\frac{1}{\delta} + \log\log\frac{1}{\varepsilon} + \log\log\frac{1}{\delta'}\right)\right),$$

which indicates that the space complexity of the algorithm is $\mathrm{poly}(\frac{1}{\delta}, \frac{1}{\delta'}, \frac{1}{\varepsilon}, \log\log n)$.

Since δ' and δ both refer to some small probability, it is acceptable to mix them up.

Exercise 6 (Bonus 10 points, graded by Zelin Li)
Recall that in the class (see Lecture note 7), we have seen one algorithm based on dimension reduction for solving (c, r)-ANN problem.
Let $0 < p \leq \frac{1}{2}$. Prove that for any $\boldsymbol{x}, \boldsymbol{y} \in \{0,1\}^d$, it holds that

$$\Pr[(U\boldsymbol{x})_i \neq (U\boldsymbol{y})_i] = \frac{1}{2}\left(1 - (1-2p)^{\mathrm{Ham}(\boldsymbol{x},\boldsymbol{y})}\right),$$

where U is a $k \times d$ random matrix such that the entries are independently and identically distributed (i.i.d.)

as follows:

$$u_{ij} = \begin{cases} 1 & \text{with probability } p, \\ 0 & \text{with probability } 1-p, \end{cases}$$

and all the calculations are in the finite field GF(2) (i.e., addition and multiplication are always modulo 2).

Hint: You may consider to use the following fact: Let $\boldsymbol{w} \in \{0,1\}^d$ be a random vector such that all entries w_i's are i.i.d. and $\Pr[w_i = 1] = \Pr[w_i = 0] = \frac{1}{2}$ for each $i \le d$. Then $\Pr\left[\boldsymbol{w}^\top \boldsymbol{x} \neq \boldsymbol{w}^\top \boldsymbol{y}\right] = \frac{1}{2}$ if $\boldsymbol{x} \neq \boldsymbol{y}$.

Proof Method 1 (by using the hint). Obviously we have $\Pr[(U\boldsymbol{x})_i \neq (U\boldsymbol{y})_i] = 0$ if $\boldsymbol{x} = \boldsymbol{y}$, so we only need to prove the case when $\boldsymbol{x} \neq \boldsymbol{y}$.
Let $\boldsymbol{u}_i^\top$ be the i-th row of U. We construct two i.i.d. random variables as follows:

$$u'_{ij} = \begin{cases} 1 & \text{with probability } \frac{1}{2}, \\ 0 & \text{with probability } \frac{1}{2}, \end{cases}$$

$$a_{ij} = \begin{cases} 1 & \text{with probability } 2p, \\ 0 & \text{with probability } 1-2p. \end{cases}$$

Note that $u_{ij} = u'_{ij} a_{ij}$. Let $\boldsymbol{u_i'}^\top = (u'_{i1}, \ldots, u'_{id})$, $\boldsymbol{x}' = (a_{i1}x_1, \ldots, a_{id}x_d)^\top$ and $\boldsymbol{y}' = (a_{i1}y_1, \ldots, a_{id}y_d)^\top$. Thus we have

$$\begin{aligned} \Pr[(U\boldsymbol{x})_i \neq (U\boldsymbol{y})_i] &= \Pr\left[\boldsymbol{u}_i^\top \boldsymbol{x} \neq \boldsymbol{u}_i^\top \boldsymbol{y}\right] \\ &= \Pr\left[\boldsymbol{u_i'}^\top \boldsymbol{x}' \neq \boldsymbol{u_i'}^\top \boldsymbol{y}'\right] \\ &= \Pr\left[\boldsymbol{u_i'}^\top \boldsymbol{x}' \neq \boldsymbol{u_i'}^\top \boldsymbol{y}' \mid \boldsymbol{x}' \neq \boldsymbol{y}'\right] \cdot \Pr[\boldsymbol{x}' \neq \boldsymbol{y}'] \end{aligned}$$

According to the hint, $\Pr\left[\boldsymbol{u_i'}^\top \boldsymbol{x}' \neq \boldsymbol{u_i'}^\top \boldsymbol{y}' \mid \boldsymbol{x}' \neq \boldsymbol{y}'\right] = \frac{1}{2}$. Next we calculate the second part

$$\Pr[\boldsymbol{x}' \neq \boldsymbol{y}'] = 1 - \Pr[\boldsymbol{x}' = \boldsymbol{y}'] = 1 - (1-2p)^{\mathrm{Ham}(\boldsymbol{x},\boldsymbol{y})}.$$

The last equality holds because all different bits in $\boldsymbol{x}$ and $\boldsymbol{y}$ must be the same in $\boldsymbol{x}'$ and $\boldsymbol{y}'$ to ensure that $\boldsymbol{x}' = \boldsymbol{y}'$. Therefore,

$$\Pr[(U\boldsymbol{x})_i \neq (U\boldsymbol{y})_i] = \frac{1}{2}\left(1 - (1-2p)^{\mathrm{Ham}(\boldsymbol{x},\boldsymbol{y})}\right).$$

□

Proof Method 2. Same as the Method 1, we only need to consider the case when $\boldsymbol{x} \neq \boldsymbol{y}$.

$$\begin{aligned} \Pr[(U\boldsymbol{x})_i \neq (U\boldsymbol{y})_i] &= \Pr\left[\sum_{j=1}^{d} u_{ij}x_j \neq \sum_{j=1}^{d} u_{ij}y_j\right] \\ &= \Pr\left[\sum_{j=1}^{d} u_{ij}(x_j + y_j) \neq 0\right] \\ &= \Pr\left[\sum_{x_j \neq y_j} u_{ij} = 1\right] \\ &= \sum_{k=0}^{\lfloor \frac{\mathrm{Ham}(\boldsymbol{x},\boldsymbol{y})}{2} \rfloor} \binom{\mathrm{Ham}(\boldsymbol{x},\boldsymbol{y})}{2k+1} p^{2k+1}(1-p)^{\mathrm{Ham}(\boldsymbol{x},\boldsymbol{y})-2k-1}. \end{aligned} \tag{2}$$

Note that we have the following equalities

$$\sum_{k=0}^{\mathrm{Ham}(\boldsymbol{x},\boldsymbol{y})} \binom{n}{k} p^k (1-p)^{\mathrm{Ham}(\boldsymbol{x},\boldsymbol{y})-k} = (p + (1-p))^{\mathrm{Ham}(\boldsymbol{x},\boldsymbol{y})} = 1, \tag{3}$$

$$\sum_{k=0}^{\mathrm{Ham}(\boldsymbol{x},\boldsymbol{y})} \binom{n}{k} (-p)^k (1-p)^{\mathrm{Ham}(\boldsymbol{x},\boldsymbol{y})-k} = (1-2p)^{\mathrm{Ham}(\boldsymbol{x},\boldsymbol{y})}. \tag{4}$$

Observe that $2 \times (2) = (3) - (4)$. Thus we have

$$\Pr[(U\boldsymbol{x})_i \neq (U\boldsymbol{y})_i] = \frac{1}{2}\left(1 - (1-2p)^{\mathrm{Ham}(\boldsymbol{x},\boldsymbol{y})}\right).$$

$\square$