Randomized Algorithms.
Lecture Notes on Nearest Neighbor Search 2.
Kasper Green Larsen.

1 Nearest Neighbor Search

This lecture note assumes that the survey by Andoni and Indyk has been read as well as the first lecture note on nearest neighbor search. Recall that we were studying the following nearest neighbor search problem:

Problem. We focus on the Randomized c-approximate R-near neigbbor search (as in Definition 2.1 in the paper by Andoni and Indyk). Recall that we are given a set P of n points in $\mathbb{R}^d$ and parameters $R > 0, \delta > 0$. We are also given a distance function dist : $\mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$. For a query point q, if there is an R-near neighbor p of q (i.e. $\text{dist}(p, q) \leq R$), we must report some point p' that is a cR-near neighbor of q (i.e. $\text{dist}(p', q) \leq cR$) with probability at least $1 - \delta$.

Sensitive Hash Functions. We defined a family of hash functions $\mathcal{H}$ to be (R, cR, P_1, P_2)-sensitive for dist if it satisfies:

1. For all $p, q \in \mathbb{R}^d$ with $\text{dist}(p, q) \leq R$: $\Pr_{h \sim \mathcal{H}}[h(p) = h(q)] \geq P_1$.
2. For all $p, q \in \mathbb{R}^d$ with $\text{dist}(p, q) > cR$: $\Pr_{h \sim \mathcal{H}}[h(p) = h(q)] \leq P_2$.

Solutions. We saw a solution with query time $O(L(d + tk))$ and space $O(nd + Ln)$ words, where $L = O(n^{\lg P_1 / \lg P_2})$, t is the time to evaluate a hash function from $\mathcal{H}$ and $k = \lg_{P_2}(1/n)$.

Hamming Distance. We analysed a family of hash functions for Hamming distance when the data points are in $\{0, 1\}^d$. The family was $\mathcal{H} = \{h_i(x) = x_i \mid i \in \{1, \ldots, d\}\}$. That is, a uniform random hash function h in $\mathcal{H}$ chooses a uniform random bit position and returns the corresponding bit as the hash value. We argued that $\mathcal{H}$ was $(R, cR, 1-R/d, 1-cR/d)$-sensitive and then we proved that the exponent $\rho = \lg P_1 / \lg P_2$ satisfies $\rho \leq 1/c$.

2 Other Distance Measures

In this note, we study other distance measures than Hamming distance and derive sensitive hash functions, which yield nearest neighbor search data structures using the construction from last lecture.

2.1 ℓ_1-distance

We start by studying the distance measure most similar to Hamming distance, namely ℓ_1 distance (a.k.a. Manhattan distance). For two points $p, q \in \mathbb{R}^d$, we have $\text{dist}_{\ell_1}(p, q) = \sum_{i=1}^d |p_i - q_i|$. Given the radius R and approximation factor c, we can design a sensitive family of hash functions as follows: For a radius w to be determined, consider the (infinite) family:

$$\mathcal{H} = \left\{ h_{i,o}(x) = \left\lfloor \frac{x_i - o}{w} \right\rfloor \mid i \in \{1, \ldots, d\}, o \in [0, w) \right\}.$$

That is, to draw a hash function from $\mathcal{H}$, a coordinate i and an offset o uniformly and independently in the interval $[0, w)$. The hash value of a point x is then $\lfloor (x_i - o)/w \rfloor$.

Consider the setup where the index i in $h_{i,o}$ has been chosen, but o is still uniform random in $[0, w)$. For points p, q, notice that $h_{i,o}(p)$ and $h_{i,o}(q)$ equal precisely when $\lfloor (p_i - o)/w \rfloor = \lfloor (q_i - o)/w \rfloor$. The expression $\lfloor (x - o_i)/w \rfloor$ changes value precisely at points $x = o_i + jw$ for $j \in \mathbb{Z}$ ($\mathbb{Z}$= integers). We are thus interested in bounding the probability that a point of the form $o_i + jw$ lies between p_i and q_i. If $|p_i - q_i| \geq w$, then this happens with probability 1. Otherwise, assume wlog. that $q_i > p_i$. Notice that the first point of the form $o_i + jw$ that lies on or after p_i is uniformly at random distributed in $[p_i, p_i + w)$. Thus there is such a point in between p_i and q_i precisely with probability $|p_i - q_i|/w$. We thus have $\Pr_o[\lfloor (p_i - o_i)/w \rfloor = \lfloor (q_i - o_i)/w \rfloor] = \max\{0, 1 - |p_i - q_i|/w\}$.

Since i is chosen uniformly, we get

$$\Pr_{h \sim \mathcal{H}}[h(p) = h(q)] = \sum_{i=1}^{d} \frac{1}{d} \cdot \max\{0, 1 - |p_i - q_i|/w\}.$$

Let us first assume $\mathrm{dist}(p,q) \leq R$. If we choose $w \geq R$, then the above simplifies to:

$$\begin{aligned}
\Pr_{h \sim \mathcal{H}}[h(p) = h(q)] &= \sum_{i=1}^{d} \frac{1}{d} \cdot (1 - |p_i - q_i|/w) \\
&= 1 - \frac{\sum_{i=1}^{d} |p_i - q_i|}{dw} \\
&= 1 - \frac{\mathrm{dist}(p,q)}{dw} \\
&\geq 1 - R/(dw).
\end{aligned}$$

On the other hand, assume $\mathrm{dist}(p,q) > cR$ and assume we choose $w \geq cR$. Then we get:

$$\begin{aligned}
\Pr_{h \sim \mathcal{H}}[h(p) = h(q)] &= \sum_{i=1}^{d} \frac{1}{d} \cdot \max\{0, 1 - |p_i - q_i|/w\} \\
&= \sum_{i=1}^{d} \frac{1}{d} \cdot \left(1 - \frac{\min\{w, |p_i - q_i|\}}{w}\right) \\
&= 1 - \frac{\sum_{i=1}^{d} \min\{w, |p_i - q_i|\}}{dw}.
\end{aligned}$$

Assume first that $\forall i : |p_i - q_i| \leq w$. Then the above equals

$$\begin{aligned}
\Pr_{h \sim \mathcal{H}}[h(p) = h(q)] &= 1 - \frac{\sum_{i=1}^{d} \min\{w, |p_i - q_i|\}}{dw} \\
&= 1 - \frac{\sum_{i=1}^{d} |p_i - q_i|}{dw} \\
&= 1 - \frac{\mathrm{dist}(p,q)}{dw} \\
&\leq 1 - cR/(dw).
\end{aligned}$$

On the other hand, assume $\exists i : |p_i - q_i| > w$. Then the above is no more than:

$$\begin{aligned}
\Pr_{h \sim \mathcal{H}}[h(p) = h(q)] &= 1 - \frac{\sum_{i=1}^{d} \min\{w, |p_i - q_i|\}}{dw} \\
&\leq 1 - \frac{w}{dw} \\
&\leq 1 - cR/(dw).
\end{aligned}$$

We have thus shown $\Pr_{h \sim \mathcal{H}}[h(p) = h(q)] \leq 1 - cR/(dw)$. The family $\mathcal{H}$ is thus $(R, cR, 1 - R/(dw), 1 - cR/(dw))$-sensitive if we choose $w \geq cR$. This results in $L = O(n^\rho)$ for

$$\rho = \frac{\lg P_1}{\lg P_2} = \frac{\lg(1 - R/(dw))}{\lg(1 - cR/(dw))} \leq \frac{\lg(1 - R/(dw))}{\lg((1 - R/(dw))^c)} = 1/c.$$

Here we used the fact from last lecture that $\lg(1-a)/\lg(1-ca) \le \lg(1-a)/\lg((1-a)^c) = 1/c$ whenever $c > 1$ and $a \in [0, 1)$.

We also have $k = \lg_{P_2}(1/n) = \ln(1/n)/\ln(P_2) = \ln(1/n)/\ln(1 - cR/(dw))$. From last week, we saw that $\ln(1/n)/\ln(x)$ is increasing in x, hence we may get an upper bound by replacing $1 - cR/(dw)$ by something bigger. For that, we use $1-x \le e^{-x}$ for all x to conclude $k \le \ln(1/n)/\ln(e^{-cR/(dw)}) = \ln(1/n)/(-cR/(dw)) = dw \ln n/(cR)$.

2.2 Set Similarity Search

We will now turn to an interesting example where the data consists of sets rather than points in $\mathbb{R}^d$. More formally, let the input consist of n sets $S_1, \ldots, S_n$, all subsets of a universe U. A classic measure of set similarity is the Jaccard coefficient, which for two sets A and B is defined as:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}.$$

Thus the Jaccard coefficient lies between 0 and 1, with $J(A, B) = 1$ when A and B are identical and $J(A, B) = 0$ when A and B are disjoint. We would like to support nearest neighbor queries on sets, i.e. given a set C, find the set S_i in the data that is most similar to C. Our goal is to use the locality sensitive hashing (LSH) machinery. However, for LSH, we need a distance to satisfy that close points/sets have small distance and far away points/sets have large distance. This is the opposite of the Jaccard coefficient. Hence we define the following distance measure:

$$\mathrm{dist}_J(A, B) = 1 - J(A, B).$$

Our goal is now to define a locality sensitive hash function for this distance measure. A very elegant solution is based on so-called *min-wise independent* families of hash functions [**?**]. A family of hash functions $\mathcal{H} \subseteq U \to \mathbb{R}$ is min-wise independent if for all $x \in U, S \subset U$ with $x \notin U$, it holds that

$$\Pr_{h \sim H}[h(x) < \min_{y \in S} h(y)] = \frac{1}{|S| + 1}.$$

The above definition means that in any set, each element has exactly the same probability of receiving the smallest hash value. For now, we assume that any $h \in \mathcal{H}$ returns distinct values for all $x \in U$, i.e. $h(x) \ne h(y)$ for any $x \ne y$ and any $h \in \mathcal{H}$.

Assume that we have such a min-wise independent family of hash functions $\mathcal{H}$ and define from it a new family of hash functions $\mathcal{G}_{\mathcal{H}} \subset 2^U \to U$ as follows:

$$\mathcal{G}_{\mathcal{H}} = \left\{ g_h(A) = \operatorname*{argmin}_{x \in A} h(x) \mid h \in \mathcal{H} \right\}.$$

This definition needs a bit of explanation. The notation 2^U refers to all subsets of the universe U. So $\mathcal{G}_{\mathcal{H}}$ consists of functions that map sets $A \subseteq U$ to hash values in U. The family $\mathcal{G}_{\mathcal{H}}$ has one function g_h for each function $h \in \mathcal{H}$. To evaluate $g_h(A)$, we compute the element $x \in A$ receiving the smallest hash value when using h, and we return that element x as the hash value of A. Said simply, $g_h(A)$ returns the element in A with smallest hash value according to h.

Let us argue that $\mathcal{G}_{\mathcal{H}}$ is sensitive with respect to the distance measure dist_J. Consider two sets A, B and let $x^* = \operatorname{argmin}_{x \in A \cup B} h(x)$ be the element with smallest hash value in the union $A \cup B$. We observe that $g_h(A) = g_h(B)$ if and only if $x^* \in A \cap B$. Now define for every $x \in A \cap B$ the event E_x such that $h(x) < \min_{y \in (A \cup B) \setminus \{x\}} h(y)$, i.e. E_x is the event that x has the smallest hash value among all elements in $A \cup B$. We then have:

$$\begin{aligned}
\Pr_{g_h \sim \mathcal{G}_{\mathcal{H}}}[g_h(A) = g_h(B)] &= \Pr_{h \sim \mathcal{H}}[\operatorname*{argmin}_{x \in A \cup B} h(x) \in A \cap B] \\
&= \Pr_{h \sim \mathcal{H}}\left[\bigcup_{x \in A \cap B} E_x\right].
\end{aligned}$$

The union bound does not help us here, because we need both an upper and lower bound on this probability to give bounds on both P_1 and P_2. Fortunately, the events E_x are *disjoint*, i.e. no two of them can happen simultaneously. We can therefore sum up their probabilities:

$$\begin{aligned}
\Pr_{g_h \sim \mathcal{G}_{\mathcal{H}}}[g_h(A) = g_h(B)] &= \Pr_{h \sim \mathcal{H}}\left[\bigcup_{x \in A \cap B} E_x\right] \\
&= \sum_{x \in A \cap B} \Pr_{h \sim \mathcal{H}}[E_x] \\
&= \sum_{x \in A \cap B} \frac{1}{|(A \cup B) \setminus \{x\}| + 1} \\
&= \frac{|A \cap B|}{|A \cup B|} \\
&= J(A, B).
\end{aligned}$$

We are now ready to bound P_1 and P_2 for fixed distances R and cR. First let A, B have $\text{dist}_J(A, B) \leq R$. We see that $R \geq \text{dist}_J(A, B) = 1 - J(A, B) = 1 - \Pr_{g_h \sim \mathcal{G}_{\mathcal{H}}}[g_h(A) = g_h(B)]$, implying $\Pr_{g_h \sim \mathcal{G}_{\mathcal{H}}}[g_h(A) = g_h(B)] \geq 1 - R$. Next let A, B have $\text{dist}_J(A, B) > cR$. We then have $cR < \text{dist}_J(A, B) = 1 - J(A, B) = 1 - \Pr_{g_h \sim \mathcal{G}_{\mathcal{H}}}[g_h(A) = g_h(B)]$ implying $\Pr_{g_h \sim \mathcal{G}_{\mathcal{H}}}[g_h(A) = g_h(B)] < 1 - cR$. The family $\mathcal{G}_{\mathcal{H}}$ is thus $(R, cR, 1 - R, 1 - cR)$-sensitive.

We can now bound ρ in $L = O(n^\rho)$ as (using the rule from last week that $\lg(1 - a) / \lg(1 - ca) \leq \lg(1 - a) / \lg((1 - a)^c) = 1/c$:

$$\rho = \lg P_1 / \lg P_2 = \lg(1 - R) / \lg(1 - cR) \leq \lg(1 - R) / \lg((1 - R)^c) = 1/c.$$

We can bound k as $k = \lg_{P_2}(1/n) = \ln(1/n) / \ln(1 - cR)$. As we did last week, we notice that $\ln(1/n) / \ln x$ is increasing in x by computing the derivative as $\ln(1/n) \cdot (-1/(\ln x)^2)/x$ which we observe is positive for all $x > 0$ (since $\ln(1/n)$ is negative). Using that $1 - x \leq e^{-x}$ for all x, we thus conclude $k \leq \ln(1/n) / \ln(e^{-cR}) = \ln(n)/(cR)$.

Constructing Min-Wise Independent Families. All of the above assumed that we had a min-wise independent family of hash functions $\mathcal{H}$. Unfortunately it can be proved that such families of hash functions are impossible to construct using a small random seed. However, we can still find efficient families that are only approximately min-wise independent. This suffices to get reasonable nearest neighbor search structures.

We say that a family of hash functions $\mathcal{H}$ is ε-approximate min-wise independent, if for any $x \in U$ and any set $S \subset U$ with $x \notin S$, we have

$$\Pr_{h \sim \mathcal{H}}[h(x) < \min_{y \in S} h(y)] \in \frac{1 \pm \varepsilon}{|S| + 1}.$$

Indyk [**?**] showed that any $O(\lg(1/\varepsilon))$-wise independent family of hash functions is also ε-approximate min-wise independent. A classic example of a k-wise independent family of hash function is the following for any prime $p > U$:

$$\mathcal{H} = \left\{ h_{a_{k-1}, \ldots, a_0}(x) = \sum_{i=0}^{k-1} a_i x^i \bmod p \mid a_0, \ldots, a_{k-1} \in [p] \right\}.$$

Patrascu and Thorup [**?**] proved that the simple tabulation hash function is ε-approximate min-wise independent with $\varepsilon = O(\lg^2 n / n^{1/c})$ where c is the number of tables used for the hash function (i.e. each key in U is split into c characters of $(1/c) \lg |U|$ bits) and n is the cardinality of the set S. So the quality of the hash function depends on the set cardinalities and therefore simple tabulation hashing works best for rather large sets.

We conclude by showing how the approximation affects the sensitivity of the hash function. Notice that:

$$\begin{aligned}
\Pr_{g_h \sim \mathcal{G}_\mathcal{H}}[g_h(A) = g_h(B)] &= \sum_{x \in A \cap B} \Pr_{h \sim \mathcal{H}}[E_x] \\
&\leq \sum_{x \in A \cap B} \frac{1+\varepsilon}{|(A \cup B) \setminus \{x\}| + 1} \\
&= (1+\varepsilon) \cdot \frac{|A \cap B|}{|A \cup B|} \\
&= (1+\varepsilon) J(A, B).
\end{aligned}$$

Therefore, for $\text{dist}_J(A, B) > cR$, we have $\Pr_{g_h \sim \mathcal{G}_\mathcal{H}}[g_h(A) = g_h(B)] \leq (1+\varepsilon)J(A,B) \leq (1+\varepsilon)(1 - \text{dist}_J(A,B)) \leq (1+\varepsilon)(1-cR)$. We also have:

$$\begin{aligned}
\Pr_{g_h \sim \mathcal{G}_\mathcal{H}}[g_h(A) = g_h(B)] &= \sum_{x \in A \cap B} \Pr_{h \sim \mathcal{H}}[E_x] \\
&\geq \sum_{x \in A \cap B} \frac{1-\varepsilon}{|(A \cup B) \setminus \{x\}| + 1} \\
&= (1-\varepsilon) \cdot \frac{|A \cap B|}{|A \cup B|} \\
&= (1-\varepsilon) J(A, B).
\end{aligned}$$

This implies that for $\text{dist}_J(A,B) \leq R$, we have $\Pr_{g_h \sim \mathcal{G}_\mathcal{H}}[g_h(A) = g_h(B)] \geq (1-\varepsilon)J(A,B) = (1-\varepsilon)(1 - \text{dist}_J(A,B)) \geq (1-\varepsilon)(1-R)$. We thus conclude that $\mathcal{G}_\mathcal{H}$ is $(R, cR, (1-R)(1-\varepsilon), (1-cR)(1+\varepsilon))$-sensitive.

2.3 ℓ_2-distance

Let us finally focus on the ℓ_2 distance (a.k.a. Euclidian distance) given by $\text{dist}_{\ell_2}(p,q) = \sqrt{\sum_{i=1}^d (p_i - q_i)^2}$. This is a very natural distance measure as well, but we have saved it for last as the math is a little more tedious than the previous examples. Let w be a parameter to be fixed and consider the following (infinite) family of hash functions:

$$\mathcal{H} = \left\{ h_{o,u}(x) = \left\lfloor \frac{\langle x, u \rangle - o}{w} \right\rfloor \mid o \in [0, w), u \in \mathbb{R}^d \right\}.$$

To draw a hash function $h_{o,u}$ from $\mathcal{H}$, we pick the offset o uniformly in $[0,w)$ and we pick the vector u such that each coordinate is independently $\mathcal{N}(0,1)$ distributed.

Consider two points p, q and two fixed values of $\langle p, u \rangle$ and $\langle q, u \rangle$, i.e. consider what happens once u has been chosen but o is still uniform random in $[0,w)$. Similarly to the analysis for ℓ_1-distance, we get that $\lfloor \frac{\langle p,u \rangle - o}{w} \rfloor = \lfloor \frac{\langle q,u \rangle - o}{w} \rfloor$ with probability $\max\{0, 1 - |\langle p,u \rangle - \langle q,u \rangle|/w\}$. By linearity, we may write:

$$\max\{0, 1 - |\langle p,u \rangle - \langle q,u \rangle|/w\} = \max\{0, 1 - |\langle p - q, u \rangle / w|\}.$$

The above probability holds for any fixed u. So let us consider the distribution of $\langle p - q, u \rangle$. Using the rule that for two independent Gaussians $X \sim \mathcal{N}(0, \sigma_x^2), Y \sim \mathcal{N}(0, \sigma_y^2)$ we have $aX + bY \sim \mathcal{N}(0, a^2\sigma_x^2 + b^2\sigma_y^2)$ we conclude that

$$\langle p - q, u \rangle = \sum_{i=1}^d (p_i - q_i) u_i \sim \mathcal{N}(0, \sum_i (p_i - q_i)^2) \sim \mathcal{N}(0, \text{dist}(p,q)^2) \sim \text{dist}(p,q)\mathcal{N}(0,1).$$

Therefore, we also have $\langle p - q, u \rangle / w \sim (\text{dist}(p,q)/w)\mathcal{N}(0,1)$. The probability density function f of the $\mathcal{N}(0,1)$ distribution is $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$. Thus the probability that $h(p) = h(q)$ equals:

$$\Pr[h(p) = h(q)] = \int_{x=-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \max\{0, 1 - (\text{dist}(p,q)/w)|x|\} dx.$$

This is at least:

$$\begin{aligned}\Pr[h(p)=h(q)] &\geq \\ \int_{x=-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \left(1-(\operatorname{dist}(p,q)/w)|x|\right) dx &= \\ \int_{x=-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx - \int_{x=-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} (\operatorname{dist}(p,q)/w)|x| dx.\end{aligned}$$

The first of these two integrals is simply the integral of the probability density function of $\mathcal{N}(0,1)$ over its full support, hence it equals 1. We use symmetry to get rid of the absolute value in the second integral and move constant factors outside:

$$\begin{aligned}\Pr[h(p)=h(q)] &\geq \\ 1 - 2\cdot(\operatorname{dist}(p,q)/w)\cdot\frac{1}{\sqrt{2\pi}}\cdot\int_{x=0}^{\infty} e^{-x^2/2} x\ dx.\end{aligned}$$

Now observe that $\int e^{-x^2/2} x\ dx = -e^{-x^2/2} + C$ for any constant C. This can be verified simply by computing the derivative of $-e^{-x^2/2}+C$. Observing that $-e^{-x^2/2} \to 0$ as $x \to \infty$ and $-e^{-x^2/2}$ equals -1 for $x=0$ we get:

$$\Pr[h(p)=h(q)] \quad \geq \quad 1 - 2\cdot(\operatorname{dist}(p,q)/w)\cdot\frac{1}{\sqrt{2\pi}}.$$

For p,q with $\operatorname{dist}(p,q) \leq R$, this is at least $1-\sqrt{2/\pi}(R/w)$. We also need an upper bound on $\Pr[h(p)=h(q)]$. Here we see that:

$$\begin{aligned}\Pr[h(p)=h(q)] &= \\ \int_{x=-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \max\{0, 1-(\operatorname{dist}(p,q)/w)|x|\} dx &\leq \\ \int_{x=-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \left(1-(\operatorname{dist}(p,q)/w)|x|\right) dx + 2\cdot\int_{x=w/\operatorname{dist}(p,q)}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} (\operatorname{dist}(p,q)/w) x\ dx &= \\ 1-\sqrt{2/\pi}(\operatorname{dist}(p,q)/w) + 2\cdot\frac{1}{\sqrt{2\pi}}\cdot(\operatorname{dist}(p,q)/w)\int_{x=w/\operatorname{dist}(p,q)}^{\infty} e^{-x^2/2} x\ dx &= \\ 1-\sqrt{2/\pi}(\operatorname{dist}(p,q)/w) + \sqrt{2/\pi}(\operatorname{dist}(p,q)/w) e^{-(w/\operatorname{dist}(p,q))^2/2} &= \\ 1-\sqrt{2/\pi}(\operatorname{dist}(p,q)/w)\left(1-e^{-(w/\operatorname{dist}(p,q))^2/2}\right).&\end{aligned}$$

Let us for simplicity assume $w \gg \operatorname{dist}(p,q)$ and let us just ignore the term $\left(1-e^{-(w/\operatorname{dist}(p,q))^2/2}\right)$. Under that simplification, we get for p,q with $\operatorname{dist}(p,q) > cR$ that $\Pr[h(p)=h(q)] \leq 1-\sqrt{2/\pi}(cR/w)$. We can now use the exact same calculations as last week ($\lg(1-a)/\lg(1-ac) \leq \lg(1-a)/\lg((1-a)^c) = 1/c$) to show that the parameter ρ in $L = O(n^\rho)$ satisfies:

$$\begin{aligned}\rho = \lg(P_1)/\lg(P_2) = \lg(1-\sqrt{2/\pi}(R/w))/\lg(1-\sqrt{2/\pi}(cR/w)) &\leq \\ \lg(1-\sqrt{2/\pi}(R/w))/\lg\left((1-\sqrt{2/\pi}(R/w))^c\right) = 1/c.&\end{aligned}$$

Note that we ignored the term $e^{-(w/\operatorname{dist}(p,q))^2/2}$ for simplicity, so the calculations are not completely correct. If we had not made the simplication, we could have concluded:

$$\rho \leq \frac{1}{c\left(1-e^{-(w/\operatorname{dist}(p,q))^2/2}\right)}.$$

Thus in choosing w, we actually need to have some estimate on the largest value of $\mathrm{dist}(p,q)$ that we will encounter. Let us finally bound k. We will do it for the simplication with $P_2 = 1-\sqrt{2/\pi}(cR/w)$, i.e. ignoring the term $\left(1 - e^{-(w/\,\mathrm{dist}(p,q))^2/2}\right)$. We have $k = \lg_{P_2}(1/n) = \ln(1/n)/\ln(P_2) = \ln(1/n)/\ln(1-\sqrt{2/\pi}(cR/w))$. Using that $1 - x \le e^{-x}$ for all x, we thus conclude $k \le \ln(1/n)/\ln\left(e^{-\sqrt{2/\pi}(cR/w)}\right) = \ln n/(\sqrt{2/\pi}cR/w) = \sqrt{\pi/2}(w\ln n)/(cR) = O(w\ln n/(cR))$. This means that we should also be careful not to choose w too large.