1(a)

For each dimension i of $\boldsymbol{y}$, y_i must lie in the range $(x_i - \epsilon, x_i + \epsilon)$ to satisfy $\max_m |x_m - y_m| \leq \epsilon$. For each dimension this interval has total length 2ϵ. So the probability of this occurring is $(2\epsilon)^M$.

1 (b)

We have that the pdf of the absolute difference of two standard normal distributions is $2x - x$ for $0 \leq x < 1$. For the probability $x_i - y_i \leq \epsilon$ in one dimension we have that $\int_0^\epsilon 2 - 2x dx = 2\epsilon - 2\epsilon^2$. In M dimensions, the probability would be $(2\epsilon - 2\epsilon^2)^M$ and since $\epsilon > 0$, we have that $(2\epsilon - 2\epsilon^2)^M \leq (2\epsilon)^M$.

1 (c)

We have that

$$||\boldsymbol{x} - \boldsymbol{y}|| = \sqrt{(x_1 - y_1)^2 + \ldots + (x_n - y_n)^2}$$

$||\boldsymbol{x} - \boldsymbol{y}||^2 = (x_1 - y_1)^2 + \ldots + (x_n - y_n)^2 \geq (\max_m |x_m - y_m|)^2$ since the max term is just one of the terms on the left and we are adding strictly positive numbers on. Taking the square root of both sides, this implies

$$||\boldsymbol{x} - \boldsymbol{y}|| \geq \max_m |x_m - y_m|$$

We have $\mathrm{P}(\max_m |x_m - y_m| \leq \epsilon) \leq p$, since we know $||\boldsymbol{x} - \boldsymbol{y}|| \geq \max_m |x_m - y_m|$, the probability of something larger than $\max_m |x_m - y_m|$ being less than ϵ must be even smaller than p since the larger something is, the less likely it is to fit within a certain range.

1 (d)

We are looking for the lower bound on the number of points needed to guarantee the nearest neighbor of a point $\boldsymbol{x}$ will be within a radius ϵ within it with probability at least $1 - \delta$. We consider the complement, what is required for the nearest neighbor not to be within a radius ϵ. This quantity satisfies $\prod_{i=1}^N P(||\boldsymbol{x} - \boldsymbol{y}_i|| > \epsilon) \leq (1-p)^N \leq \delta$. So we have $(1 - (2\epsilon)^M)^N \leq \delta$. Solving for N, we get $N \geq \frac{\log \delta}{\log(1-(2\epsilon)^M)}$ as a lower bound, where the sign flips because the log quantity is between 0 and 1 and takes a negative value.

1 (e)

This tells us that in high dimensional spaces, points get spread more and more apart. Where M gets larger, we need more and more points as a lower bound

to guarantee some probability of having another point within a certain radius of another. This is because as M gets larger, the denominator of the inequality gets smaller, making the overall right hand side of the equation larger. Since points are further and further apart in higher dimensions, it reduces the ability of HAC to correctly group points together into clusters because as points get further apart, the ability to distinguish clusters based on distances decreases with the metrics.