Metric Embedding Theory and its Algorithms (67720) – Exercise 1

Mike Greenbaum 211747639

November 2021

1. (a) Given a metric Space $X = \{x_1, x_2, x_3\}$ and metric d_X, we define the tree embedding to be $V = \{x_1, x_2, x_3, v\}$ where $E = \{\{v, v_1\}, \{v, v_2\}, \{v, v_3\}\}$ and $w : E \to \mathbb{R}^+$ defined in the following way:

$$w(\{v, v_1\}) = \frac{d_X(x_1, x_2) + d_X(x_1, x_3) - d_X(x_2, x_3)}{2}$$
$$w(\{v, v_2\}) = \frac{d_X(x_1, x_2) + d_X(x_2, x_3) - d_X(x_1, x_3)}{2}$$
$$w(\{v, v_3\}) = \frac{d_X(x_2, x_3) + d_X(x_1, x_3) - d_X(x_1, x_2)}{2}$$

The embedding is $f(x_i) = v_i, \forall i \in [3]$.
Due to X being a metric space, the triangle inequality holds and all the expressions are non negative and therefore w is well defined. Notice that the graph is indeed a tree.
Notice that the only path between v_1, v_2 is v_1, v_2 and therefore the following holds:

$$d_W(v_1, v_2) \overset{\text{def}}{=} d_W(v_1, v) + d_W(v, v_2) = w(\{v, v_1\}) + w(\{v, v_2\}) \overset{\text{def}}{=}$$
$$\frac{d_X(x_1, x_2) + d_X(x_1, x_3) - d_X(x_2, x_3)}{2} + \frac{d_X(x_1, x_2) + d_X(x_2, x_3) - d_X(x_1, x_3)}{2} =$$
$$\frac{d_X(x_1, x_2) + d_X(x_1, x_3) - d_X(x_2, x_3) + d_X(x_1, x_2) + d_X(x_2, x_3) - d_X(x_1, x_3)}{2} =$$
$$\frac{2d_X(x_1, x_2)}{2} = d_X(x_1, x_2)$$

Therefore we got that $d_W(v_1, v_2) = d_X(x_1, x_2)$, due to the symmetry in W definition and the graph's structure, we can deduce that $\forall i \neq j \in [3], d_W(v_i, v_j) = d_X(x_i, x_j)$, therefore the construction we gave embeds isometrically in a tree metric.
Therefore, we showed that for every metric Space X with 3 elements, we can construct an isometric embedding to a tree metric.

(b) Given a metric Space $X = \{x_1, x_2, x_3\}$, define the embedding in the following way:

$$f(x_1) = (0, 0), f(x_2) = (0, d_X(x_1, x_2))$$
$$f(x_3) = (\frac{d_X(x_2, x_3) + d_X(x_1, x_3) - d_X(x_1, x_2)}{2}, \frac{d_X(x_1, x_2) + d_X(x_1, x_3) - d_X(x_2, x_3)}{2})$$

Notice that:

$$d_{l_1^2}(f(x_1), f(x_2)) = \|(0, 0) - (0, d_X(x_1, x_2))\|_1 = \|(0, d_X(x_1, x_2))\|_1 = d_X(x_1, x_2)$$

Now notice that:

$$d_{l_1^2}(f(x_1), f(x_3)) = \|(0, 0) - (\frac{d_X(x_2, x_3) + d_X(x_1, x_3) - d_X(x_1, x_2)}{2}, \frac{d_X(x_1, x_2) + d_X(x_1, x_3) - d_X(x_2, x_3)}{2})\|_1 =$$
$$\|(\frac{d_X(x_2, x_3) + d_X(x_1, x_3) - d_X(x_1, x_2)}{2}, \frac{d_X(x_1, x_2) + d_X(x_1, x_3) - d_X(x_2, x_3)}{2})\|_1 =$$
$$|\frac{d_X(x_2, x_3) + d_X(x_1, x_3) - d_X(x_1, x_2)}{2}| + |\frac{d_X(x_1, x_2) + d_X(x_1, x_3) - d_X(x_2, x_3)}{2}| =$$
$$\frac{d_X(x_2, x_3) + d_X(x_1, x_3) - d_X(x_1, x_2)}{2} + \frac{d_X(x_1, x_2) + d_X(x_1, x_3) - d_X(x_2, x_3)}{2} = d_X(x_1, x_3)$$

And finally notice that:

$$
\begin{aligned}
&d_{l_1^2}(f(x_2), f(x_3)) = \\
&\|(0, d_X(x_1,x_2)) - (\frac{d_X(x_2,x_3)+d_X(x_1,x_3)-d_X(x_1,x_2)}{2}, \frac{d_X(x_1,x_2)+d_X(x_1,x_3)-d_X(x_2,x_3)}{2})\|_1 = \\
&\|(-\frac{d_X(x_2,x_3)+d_X(x_1,x_3)-d_X(x_1,x_2)}{2}, \frac{d_X(x_1,x_2)-d_X(x_1,x_3)+d_X(x_2,x_3)}{2})\|_1 = \\
&|-\frac{d_X(x_2,x_3)+d_X(x_1,x_3)-d_X(x_1,x_2)}{2}| + |\frac{d_X(x_1,x_2)-d_X(x_1,x_3)+d_X(x_2,x_3)}{2}| = \\
&\frac{d_X(x_2,x_3)+d_X(x_1,x_3)-d_X(x_1,x_2)}{2} + \frac{d_X(x_1,x_2)-d_X(x_1,x_3)+d_X(x_2,x_3)}{2} = d_X(x_2,x_3)
\end{aligned}
$$

So we showed an embedding f s.t. $d_{l_1^2}(f(x_i), f(x_j)) = d_X(x_i,x_j), \forall i \neq j \in [3]$ and therefore we showed that any 3-point metric embeds isometrically in l_1^2.

(c) i. given $G = (V,E)$, Assume wlog that the distance on each edge in the graph is 1, else scale the answer by that scalar. We define the following embedding:

$$f(w) = (0,0), f(x) = (0,1), f(y) = (\sqrt{\frac{3}{4}}, -\frac{1}{2}), f(z) = (-\sqrt{\frac{3}{4}}, -\frac{1}{2})$$

Notice that:

$$
\begin{aligned}
&d_{l_2^2}(x,w) = d_{l_2^2}(y,w) = d_{l_2^2}(z,w) = 1 \\
&d_{l_2^2}(x,y) = d_{l_2^2}(x,z) = \sqrt{(\sqrt{\frac{3}{4}})^2 + (1+\frac{1}{2})^2} = \sqrt{3} \\
&d_{l_2^2}(y,z) = 2 \cdot \sqrt{\frac{3}{4}} = \sqrt{3}
\end{aligned}
$$

Therefore, by definition: $contr(f) = \frac{2}{\sqrt{3}}, expans(f) = 1$, Therefore, $dist(f) = \frac{2}{\sqrt{3}} \cdot 1 = \frac{2}{\sqrt{3}}$, as required.

ii. As said in the hint, notice that $\frac{\partial \|x-w\|_2^2 + \|y-w\|_2^2 + \|z-w\|_2^2}{\partial w} = 2(x-w)^T + 2(y-w)^T + 2(z-w)$.
Notice that it is zero only at $w = \frac{x+y+z}{3}$.
Notice that the function is a summation of convex functions and therefore $w = \frac{x+y+z}{3}$ is the minimum of the function, Therefore:

$$
\begin{aligned}
&3 \cdot (\|x-w\|_2^2 + \|y-w\|_2^2 + \|z-w\|_2^2) \\
&\geq 3 \cdot (\|x - \frac{x+y+z}{3}\|_2^2 + \|y - \frac{x+y+z}{3}\|_2^2) + \|z - \frac{x+y+z}{3}\|_2^2) \\
&= 3 \cdot (\|\frac{2x-y-z}{3}\|_2^2 + \|\frac{2y-x-z}{3}\|_2^2 + \|\frac{2z-x-y}{3}\|_2^2) \\
&= \frac{3}{9} \cdot (\sum_i 4x_i^2 - 4x_iy_i - 4x_iz_i + y_i^2 + 2y_iz_i + z_i^2 \\
&+ \sum_i 4y_i^2 - 4y_ix_i - 4y_iz_i + x_i^2 + 2x_iz_i + z_i^2 \\
&+ \sum_i 4z_i^2 - 4z_ix_i - 4z_iy_i + x_i^2 + 2x_iy_i + y_i^2) \\
&= \frac{1}{3} \cdot \sum_i 6x_i^2 - 6x_iy_i - 6x_iz_i + 6y_i^2 - 6y_iz_i + 6z_i^2 \\
&= \sum_i 2x_i^2 - 2x_iy_i - 2x_iz_i + 2y_i^2 - 2y_iz_i + 2z_i^2 \\
&= \sum_i (x_i^2 - 2x_iy_i + y_i^2) + (y_i^2 - 2y_iz_i + z_i^2) + (x_i^2 - 2x_iz_i + z_i^2) \\
&= \sum_i (x_i - y_i)^2 + (y_i - z_i)^2 + (x_i - z_i)^2 \\
&= \sum_i (x_i - y_i)^2 + \sum_i (y_i - z_i)^2 + \sum_i (x_i - z_i)^2 \\
&= \|x-y\|_2^2 + \|y-z\|_2^2 + \|x-z\|_2^2
\end{aligned}
$$

iii. let f be an embedding of G into $l2$. wlog assume that $contr(f) = 1$, therefore $\alpha \stackrel{\text{def}}{=} expans(f) = dist(f)$. Notice that, $d_X^2(a,b) \le \|f(a) - f(b)\|_2^2 \le \alpha^2 \cdot d_X^2(a,b)$, therefore:

$$\begin{aligned}
12 = 2^2 + 2^2 + 2^2 &= d_X^2(x,y) + d_X^2(y,z) + d_X^2(x,z) \\
&\le \|f(x) - f(y)\|_2^2 + \|f(y) - f(z)\|_2^2 + \|f(x) - f(z)\|_2^2 \\
&\le 3 \cdot (\|f(x) - f(w)\|_2^2 + \|f(y) - f(w)\|_2^2 + \|f(z) - f(w)\|_2^2) \\
&\le 3 \cdot (\alpha^2 \cdot d_X^2(x,w) + \alpha^2 \cdot d_X^2(y,w) + \alpha^2 \cdot d_X^2(x,w)) \\
&= 3 \cdot (\alpha^2 + \alpha^2 + \alpha^2) = 9 \cdot \alpha^2
\end{aligned}$$

$\implies 12 \le 9\alpha^2 \implies \frac{4}{3} \le \alpha^2 \implies \frac{2}{\sqrt{3}} \le \alpha$, So we showed that $\frac{2}{\sqrt{3}} \le \alpha \stackrel{\text{def}}{=} dist(f)$, as required.

2. (a) We will assume that $\log_2(n) \in \mathbb{Z}$, otherwise we add points to the space to get a power of 2.
Let $X = x_1, \ldots, x_n$. define $CUBE = \{0, d_X(x_1, x_2)\}^{\log_2(n)}$.
Notice that CUBE has $2^{\log_2(n)} = n$ elements ans we will denote $CUBE = \{v_1, \ldots, v_n\}$.
We define the embedding $f(x_i) = v_i$.
Notice that $\forall i \ne j$, v_i, v_j differ at least in one coordinate and each coordinate they differ, one has 0 and the other has $d_X(x_1, x_2)$, therefore $\|v_i - v_j\|_\infty = d_X(x_1, x_2)$. We know that the distance between each 2 points in X is the same, therefore $\forall i \ne j$, $d_X(x_i, x_j) = d_X(x_1, x_2)$.
Therefore, $\forall i \ne j$, $d_\infty(f(x_i), f(x_j) = \|v_i - v_j\|_\infty = d_X(x_1, x_2) = d_X(x_i, x_j)$.
So we showed that $\forall i \ne j, d_\infty(f(x_i), f(x_j) = d_X(x_i, x_j)$, therefore f is an isometry from X to $l_\infty^{\log_2(n)}$.

(b) define $g(\varepsilon) = \frac{(1+\varepsilon)^p - 1}{1 + (1+\varepsilon)^p}$ define $z = \frac{1 + (1+\varepsilon)^p - 2}{2} \cdot d_X(x_1, x_2)$. We know that $\frac{2\varepsilon^2 \cdot (2\ln(n) + \ln(2))}{\ln(n) \cdot g^2(\varepsilon) \cdot \log_2(e)}$ is bounded both for $0 \le \varepsilon \le 1$ and $n \in \mathbb{N}$, assume m is the bound. define $a \ge m$ define $k \stackrel{\text{def}}{=} \frac{a \cdot \log_2(n)}{\varepsilon^2}$ and define c s.t. $k \cdot \frac{1}{2} \cdot c^p = d_X^p(x_i, x_j) + z$
We will assume that $\log_2(n) \in \mathbb{Z}$, otherwise we add points to the space to get a power of 2. Let $X = x_1, \ldots, x_n$. define $CUBE = \{0, c\}^k$.
Notice that CUBE has 2^k elements ans we will denote $CUBE = \{v_1, \ldots, v_{2^k}\}$. We will define $f(x_i) = v_j$ where $j \sim Uniform\{1, \ldots, 2^k\}$. We will define the Random variable $D_{i,j,l} \stackrel{\text{def}}{=} |f(x_i)_l - f(x_j)_l|^p$ and $D_{i,j} = \sum_{l=1}^{2^k} D_{i,j,l} = \sum_{l=1}^{2^k} |f(x_i)_l - f(x_j)_l|^p = \|f(x_i) - f(x_j)\|_p^p$.
Notice that they agree on coordinates with probability $\frac{1}{2}$ and disagree with probability $\frac{1}{2}$, Therefore

$$\begin{aligned}
\mathbb{E}(D_{i,j}) &= \mathbb{E}(\sum_{l=1}^k D_{i,j,l}) = \mathbb{E}(\sum_{l=1}^k |f(x_i)_l - f(x_j)_l|^p) = \sum_{l=1}^k \mathbb{E}(|f(x_i)_l - f(x_j)_l|^p) \\
&= \sum_{l=1}^k (0 \cdot \frac{1}{2} + c^p \cdot \frac{1}{2}) = k \cdot c^p \cdot \frac{1}{2} \stackrel{\text{def}}{=} d_X^p(x_i, x_j) + z
\end{aligned}$$

Notice that due to Hoeffding

$$\begin{aligned}
&P(|\|f(x_i) - f(x_j)\|_p^p - \mathbb{E}(D_{i,j})| > g(\varepsilon) \cdot \mathbb{E}(D_{i,j})) = P(|D_{i,j} - \mathbb{E}(D_{i,j})| > g(\varepsilon) \cdot \mathbb{E}(D_{i,j})) \\
&\le 2 \cdot \exp\{-2 \frac{\cdot (g(\varepsilon) \cdot \mathbb{E}(D_{i,j})^2}{\sum_{l=1}^k c^{2p}}\} = 2 \cdot \exp\{-2 \frac{\cdot (g(\varepsilon) \cdot k \cdot c^p \cdot \frac{1}{2})^2}{k \cdot c^{2p}}\} \\
&= 2 \cdot \exp\{-2 \frac{\cdot g^2(\varepsilon) \cdot k^2 \cdot \frac{1}{4} \cdot c^{2p}}{k \cdot c^{2p}}\} = 2 \cdot \exp\{-2 \cdot g^2(\varepsilon) \cdot \frac{1}{4} \cdot k\} = 2 \cdot \exp\{g^2(\varepsilon) \cdot \frac{-1}{2} \cdot k\}
\end{aligned}$$

Now notice that

$$\begin{aligned}
&n^2 \cdot 2 \cdot \exp\{g^2(\varepsilon) \cdot \frac{-1}{2} \cdot k\} < 1 \iff n^2 \cdot 2 \cdot 2^{g^2(\varepsilon) \cdot \frac{-1}{2} \cdot k \cdot \log_2(e)} < 1 \iff n^2 \cdot 2 \cdot 2^{g^2(\varepsilon) \cdot \frac{-1}{2} \cdot \frac{a \log_2(n)}{\varepsilon^2} \cdot \log_2(e)} < 1 \\
&\iff n^2 \cdot 2 \cdot n^{g^2(\varepsilon) \cdot \frac{-1}{2} \cdot \frac{a}{\varepsilon^2} \cdot \log_2(e)} < 1 \iff n^{2 - g^2(\varepsilon) \cdot \frac{1}{2} \cdot \frac{a}{\varepsilon^2} \cdot \log_2(e)} < \frac{1}{2} \\
&\iff \ln(n) \cdot \{2 - g^2(\varepsilon) \cdot \frac{1}{2} \cdot \frac{a}{\varepsilon^2} \cdot \log_2(e)\} < -\ln(2) \iff \ln(n) \cdot 2 + \ln(2) < \ln(n) \cdot g^2(\varepsilon) \cdot \frac{1}{2} \cdot \frac{a}{\varepsilon^2} \cdot \log_2(e) \\
&\iff \frac{2\ln(n) + \ln(2)}{\frac{\ln(n) \cdot g^2(\varepsilon) \cdot \log_2(e)}{2\varepsilon^2}} < a \iff \frac{2\varepsilon^2 \cdot (2\ln(n) + \ln(2))}{\ln(n) \cdot g^2(\varepsilon) \cdot \log_2(e)} < a
\end{aligned}$$

Therefore, due to our choice of a we get that $n^2 \cdot 2 \cdot \exp\{g^2(\varepsilon) \cdot \frac{-1}{2} \cdot k\} < 1$, Therefore

$$P(\exists i \neq j, |\|f(x_i) - f(x_j)\|_p^p - \mathbb{E}(D_{i,j})| > g(\varepsilon) \cdot \mathbb{E}(D_{i,j})) \leq \sum_{i \neq j} P(|\|f(x_i) - f(x_j)\|_p^p - \mathbb{E}(D_{i,j})| > g(\varepsilon) \cdot \mathbb{E}(D_{i,j}))$$
$$\leq \sum_{i \neq j} 2 \cdot \exp\{g^2(\varepsilon) \cdot \frac{-1}{2} \cdot k\} \leq n^2 \cdot 2 \cdot \exp\{g^2(\varepsilon) \cdot \frac{-1}{2} \cdot k\} < 1$$

Notice that

$$\begin{aligned}
&(1 - g(\varepsilon)) \cdot (d_X^p(x_i, x_j) + z) = (1 - \frac{(1+\varepsilon)^p - 1}{1 + (1+\varepsilon)^p}) \cdot (d_X^p(x_i, x_j) + \frac{1 + (1+\varepsilon)^p - 2}{2} \cdot d_X(x_1, x_2)) \\
&= (\frac{1 + (1+\varepsilon)^p - (1+\varepsilon)^p + 1}{1 + (1+\varepsilon)^p}) \cdot (\frac{1 + (1+\varepsilon)^p}{2} \cdot d_X(x_1, x_2)) \\
&= (\frac{2}{1 + (1+\varepsilon)^p}) \cdot (\frac{1 + (1+\varepsilon)^p}{2} \cdot d_X(x_1, x_2)) = d_X(x_1, x_2) = d_X(x_i, x_j)
\end{aligned}$$

Furthermore, notice that:

$$\begin{aligned}
&(1 + g(\varepsilon)) \cdot (d_X^p(x_i, x_j) + z) = (1 + \frac{(1+\varepsilon)^p - 1}{1 + (1+\varepsilon)^p}) \cdot (d_X^p(x_i, x_j) + \frac{1 + (1+\varepsilon)^p - 2}{2} \cdot d_X(x_1, x_2)) \\
&= (\frac{1 + (1+\varepsilon)^p (1+\varepsilon)^p - 1}{1 + (1+\varepsilon)^p}) \cdot (\frac{1 + (1+\varepsilon)^p}{2} \cdot d_X(x_1, x_2)) \\
&= (\frac{2(1+\varepsilon)^p}{1 + (1+\varepsilon)^p}) \cdot (\frac{1 + (1+\varepsilon)^p}{2} \cdot d_X(x_1, x_2)) = (1+\varepsilon)^p \cdot d_X(x_1, x_2) = d_X(x_i, x_j) \cdot (1+\varepsilon)^p
\end{aligned}$$

Therefore

$$\begin{aligned}
&|\|f(x_i) - f(x_j)\|_p^p - \mathbb{E}(D_{i,j})| \leq g(\varepsilon) \cdot \mathbb{E}(D_{i,j}) \\
&\implies |\|f(x_i) - f(x_j)\|_p^p - (d_X^p(x_i, x_j) + z)| \leq g(\varepsilon) \cdot (z + d_X^p(x_i, x_j)) \\
&\implies (1 - g(\varepsilon)) \cdot (d_X^p(x_i, x_j) + z) \leq \|f(x_i) - f(x_j)\|_p^p \leq (1 + g(\varepsilon)) \cdot (d_X^p(x_i, x_j) + z) \\
&\implies d_X^p(x_i, x_j) \leq \|f(x_i) - f(x_j)\|_p^p \leq d_X^p(x_i, x_j) \cdot (1+\varepsilon)^p \\
&\implies d_X(x_i, x_j) \leq \|f(x_i) - f(x_j)\|_p \leq d_X(x_i, x_j) \cdot (1+\varepsilon)
\end{aligned}$$

Notice that we showed that $P(\exists i \neq j, |\|f(x_i) - f(x_j)\|_p^p - \mathbb{E}(D_{i,j})| > g(\varepsilon) \cdot \mathbb{E}(D_{i,j})) < 1$, therefore $\exists f$ s.t. $\forall i \neq j, |\|f(x_i) - f(x_j)\|_p^p - \mathbb{E}(D_{i,j})| \leq g(\varepsilon) \cdot \mathbb{E}(D_{i,j})$.
Due to the last remark, the following is also true $d_X(x_i, x_j) \leq \|f(x_i) - f(x_j)\|_p \leq (1+\varepsilon) \cdot d_X^p(x_i, x_j)$.
Therefore, we found an embedding f s.t. $dist(f) \leq 1 + \varepsilon$ and $k = O(\frac{\log_2(n)}{\varepsilon^2})$. Notice that we have a dependence on $\frac{1}{g^2(\varepsilon)} = \frac{((1+\varepsilon)^p + 1)^2}{((1+\varepsilon)^p - 1)^2}$.

3. (a) wlog we will assume that $contr(f) = 1$, i.e. $expans(f) = \alpha$, therefore, $d_X(x, y) \leq d_Y(f(x), f(y)) \leq \alpha \cdot d_X(x, y)$.
First, we will prove a helper lemma: $f(B_{\frac{r}{\alpha}}(x)) \subseteq B_r(f(x))$. Let $z \in f(B_{\frac{r}{\alpha}}(x))$, therefore $\exists y \in B_{\frac{r}{\alpha}}(x)$ s.t. $f(y) = z$, Notice that $d_X(x, y) \leq \frac{r}{\alpha}$, therefore $d_Y(f(x), z) = d_Y(f(x), f(y)) \leq \alpha \cdot d_X(x, y) \leq \alpha \cdot \frac{r}{\alpha} = r$, therefore by definition $z \in B_r(f(x))$, thus, $f(B_{\frac{r}{\alpha}}(x)) \subseteq B_r(f(x))$.
Furthermore, we will prove that $B_r(f(x)) \subseteq f(B_r(x))$. Let $z \in B_r(f(x))$, therefore $\exists y \in Y$ s.t. $f(y) = z$, Notice that $d_X(x, y) \leq d_Y(f(x), f(y)) = r$, therefore by definition $y \in B_r(x)$, therefore $z = f(y) \in f(B_r(x))$, thus, $B_r(f(x)) \subseteq f(B_r(x))$.
Now, we know that we can cover $B_r(f(x))$ with $2^{\dim(f(X))}$ balls of radius $\frac{r}{2}$.
Therefore we can write $\bigcup_{i=1}^{2^{\dim(f(X))}} B_{\frac{r}{2}}(f(x_i)) = B_r(f(x))$
Due to the lemmas we previously proved, we can deduce that

$$\bigcup_{i=1}^{2^{\dim(f(X))}} f(B_{\frac{r}{2\alpha}}(x_i)) \subseteq \bigcup_{i=1}^{2^{\dim(f(X))}} B_{\frac{r}{2}}(f(x_i)) = B_r(f(x)) \subseteq f(B_r(x))$$

Notice that if we show that for m balls $\bigcup_{i=1}^{m} f(B_{\frac{r}{2\alpha}}(x_i)) = f(B_r(x))$, we can deduce that due to minimallity of $\dim(f(X))$ the following holds $\dim(f(X)) \leq \log_2(m)$.

The same argument holds for the case that for m balls $\bigcup_{i=1}^{m} B_{\frac{r}{2\alpha}}(x_i) = B_r(x)$.

Notice that we know that to cover a ball of radius r in X we need at most $2^{\dim(X)}$ balls of radius $\frac{r}{2}$.

define $m_r(x, n)$ to be the amount of balls that we need to cover $B_r(x)$ with balls of size $\frac{r}{2^n}$.

Notice that due to the fact that we need $2^{\dim(X)}$ balls of radius $\frac{r}{2}$ for every r, We can cover a ball of radius $\frac{r}{2^{n-1}}$ with $2^{\dim(X)}$ balls of radius $\frac{r}{2^n}$, therefore it holds that $m_r(x, n) \leq m_r(x, n-1) \cdot 2^{\dim(X)}$ and $m_r(x, 1) \leq 2^{\dim(X)}$ by definition.

Therefore, inductively $m_r(x, n) \leq (2^{\dim(X)})^n$. Notice that in our case we want to know how many balls of radius $\frac{r}{2\alpha}$ can cover the space and therefore interested in $m_r(x, O(\log_2(\alpha)))$ due to the balls being of radius $\frac{r}{2\alpha}$ or smaller. Thus, We know that we need at most $m_r(x, O(\log_2(\alpha))) \leq (2^{\dim(X)})^{O(\log_2(\alpha))} = (O(\alpha))^{\dim(X)}$ balls.

Therefore we can conclude that $\dim(f(X)) \leq \log_2(m) \leq \log_2((O(\alpha))^{\dim(X)}) = \dim(X) \cdot O(\log(\alpha))$ as required.

(b) let $X = \{2^i \cdot e_i \mid i \in [d]\}$ and the norm be l_∞. Notice that X spans $\mathbb{R}^n$. Denote $x_i \stackrel{\text{def}}{=} 2^i \cdot e_i$, therefore $X = \{x_1, \ldots, x_d\}$. Notice that $d(x_i, x_j) = d(2^i \cdot e_i, 2^j \cdot e_j) = 2^{\max\{i,j\}}$. Now we will prove that $\dim(X) = O(1)$. let $x \in X$, therefore $\exists i \in [d]$ s.t. $x = x_i = 2^i \cdot e_i$. let $r > 0$, define $j \stackrel{\text{def}}{=} \max\{j \mid x_j \in B_r(x_i) \wedge j \neq i\}$. We will split into 3 cases:

i. If $j > i$, Then we know that $2^j \leq r$. Therefore, $B_r(x_i) = \{x_1, \ldots, x_j\}$ due to the way the distance function is defined. Notice that $2^{j-1} \leq \frac{r}{2} < 2^j$ and therefore $B_{\frac{r}{2}}(x_i) = \{x_1, \ldots, x_{j-1}\}$, and therefore $B_r(x_i) = \{x_1, \ldots, x_{j-1}\} \cup \{x_j\} = B_{\frac{r}{2}}(x_i) \cup B_{\frac{r}{2}}(x_j)$. So we showed that 2 balls can cover the space.

ii. If $j < i$, due to the way the distance function we know that $X = \{x_1, \ldots, x_i\}$ and due to our definition, $2^i \leq r < 2^{i+1}$, therefore $2^{i-1} \leq \frac{r}{2} < 2^i$. therefore $B_{\frac{r}{2}}(x_{i-1}) = \{x_1, \ldots, x_{i-1}\}$, and therefore $B_r(x_i) = \{x_1, \ldots, x_{i-1}\} \cup \{x_i\} = B_{\frac{r}{2}}(x_{i-1}) \cup B_{\frac{r}{2}}(x_i)$. So we showed that 2 balls can cover the space.

iii. Otherwise there are no items in $B_r(x_i)$ except x_i and therefore $B_r(x_i) = B_{\frac{r}{2}}(x_i)$. So we showed that 1 ball can cover the space.

Therefore we need at most 2 balls in all cases. Therefore, by definition $\dim(X) = \log_2(2) = 1 = O(1)$, as required.