CS 602 Assignment-2

Saksham Rathi
22B1003

CS 602 - Applied Algorithms: Assignment 2

Total Marks - 60

Instructions. Please try to be brief, clear, and technically precise. Use pseudo-codes to describe the algorithms. To solve the problems, one may assume that the instances are in general position, unless stated otherwise. Novelty in the answer carries marks.

Question 1[15 Marks] Let $(\mathcal{X}, \mathcal{R})$ and $(\mathcal{X}, \mathcal{R}')$ be two set systems with bounded VC dimension.

Show that $\mathsf{VC}\ \mathsf{dim}(\mathcal{R} \cup \mathcal{R}') \leq \mathsf{VC}\ \mathsf{dim}(\mathcal{R}) + \mathsf{VC}\ \mathsf{dim}(\mathcal{R}') + 1$.

Hint: Use Sauer's lemma.

Question 2[15 Marks] *Recall the definition of a metric space. We now introduce a metric called the* ***shortest-path metric*** *of a graph. Given a simple, undirected graph G with vertex set V, the distance between any two vertices $u, v \in V$ is defined as the length of the shortest path connecting u and v in G, where the path length is measured by the number of edges it contains. We assume that G is connected. As a simple example, consider the complete graph K_n, which gives the n-point* ***equilateral space****, where the distance between every pair of vertices is 1.*

Show that any tree and its corresponding shortest-path metric on the vertices can be *isometrically embedded* into l_1 (also known as Manhattan distance).

Hint: Begin by considering the embedding of trees with unit-length edges.

Question 3[15 Marks] Consider the star graph below and its corresponding shortest-path metric on the vertices (as defined in question 2).

1. Prove that this graph cannot be isometrically embedded into l_2.
2. Show that minimum distortion for embedding this graph into l_2 is $\frac{2}{\sqrt{3}}$.

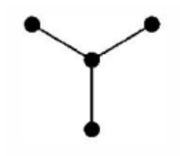


Figure 1: Star graph

Question 4[15 Marks] Let $\mathcal{U}$ be a set of size n, and let $\mathcal{X} \subseteq \mathcal{U}$ be a subset of size k. Let $\xi : \mathcal{U} \to [k]$ be a coloring of the elements of $\mathcal{U}$, chosen uniformly at random (i.e. each element of $\mathcal{U}$ is colored with one of the k colors uniformly and independently at random). Then the probability that the elements of X are colored with pairwise distinct colors is at least e^{-k}.

1b

$S = (X, R) \quad S' = (X, R') \quad S'' = (X, R \cup R')$

$\dim_{VC}(S) = \delta \quad \dim_{VC}(S') = \delta'$

Let x'' be the largest set that can be shattered by S'' i.e. $\dim_{VC}(S'') = x''$ $\quad \delta'' = \dim_{VC}(S'')$

Clearly, $|x''| \geq \max(\delta, \delta')$ as any set of size $\max(\delta, \delta')$ can be shattered by either the ranges R or R'

W.L.O.G. $\delta \geq \delta'$

Claim: S'' cannot shatter a subset of size greater than $\delta + \delta' + 1$

Proof: Say S'' can shatter a subset of size $\delta + \delta' + 2$ $\Rightarrow |R \cup R'| \geq 2^{\delta+\delta'+2}$

since $2^{x''} \leq |R \cup R'|$ if x'' is to be shattered by S''.

$|R \cup R'| \leq |R| + |R'| \leq \sum_{i=0}^{\delta} \binom{n}{i} + \sum_{i=0}^{\delta'} \binom{n}{i}$ (from union bound and sauer's lemma)

If $\delta + \delta' + 2 > n$ then $\delta'' \leq n < \delta + \delta' + 2$
$\Rightarrow \delta'' \leq \delta + \delta' + 1 \Rightarrow$ contradiction

case where $\delta + \delta' + 2 \leq n$: (X', R) and (X', R) $\rightarrow$ consider these range spaces

where $X' \subseteq X$ $\quad |X'| = \delta + \delta' + 2$

Clearly $\dim_{VC}(X', R) \leq \delta$ and $\dim_{VC}(X', R') \leq \delta'$ (because this is a subset)

$$|R \cup R'| \leq |R| + |R'| \leq \sum_{i=0}^{\delta} \binom{|X'|}{i} + \sum_{i=0}^{\delta'} \binom{|X'|}{i}$$

$$|R \cup R'| \leq \sum_{i=0}^{\delta} \binom{\delta+\delta'+2}{i} + \sum_{i=0}^{\delta'} \binom{\delta'+\delta+2}{i} = \sum_{i=0}^{\delta} \binom{\delta+\delta'+2}{i} + \sum_{i=0}^{\delta'} \binom{\delta'+\delta+2}{\delta'+\delta+2-i}$$

$$= \sum_{i=0}^{\delta} \binom{\delta+\delta'+2}{i} + \sum_{i=\delta+2}^{\delta+\delta'+2} \binom{\delta+\delta'+2}{i}$$

$$= \sum_{i=0}^{\delta+\delta'+2} \binom{\delta+\delta'+2}{i} - \binom{\delta+\delta'+2}{\delta+1} = 2^{\delta+\delta'+2} - \binom{\delta+\delta'+2}{\delta+1} \leq 2^{\delta+\delta'+2}$$

$$|R \cup R'| < 2^{\delta+\delta'+2}$$

$\Rightarrow$ A contradiction

$\Rightarrow S''$ cannot shatter of size $\delta + \delta' + 2$

$\Rightarrow \dim_{VC}(X, R \cup R') \leq \dim_{VC}(X, R) + \dim_{VC}(X, R') + 1$

2) firstly we prove that a line graph can be embedded exactly into l_1 over R^1. $G=(V,E)$ with vertices $V=\{1 \ldots n\}$
edges $E=\{(1,2)\ (2,3) \ldots [n-1,n]\}$

embedding can be defined recursively: $f(1)=0$ $f(i)=f(i-1)+d_G(i-1,i)$

Now, consider a tree $T=(V,E)$ on n vertices. This can be embedded exactly into l_1 over R^{n-1} which can be seen by induction. As the base case, when $|V|=2$, T is a line graph, which we have seen can be embedded into l_1 over R,

Tree $T_k=(V_k,E_k)$ with k vertices. Since it is a tree, we must have leaf v' and the corresponding edge (u', v'). We can remove these to obtain the tree $T_{k-1}=(V_{k-1},E_{k-1})$ and inductively embed this into l_1. $f: V_{k-1} \to R^{k-2}$ (assumption of induction)

$g: V_k \to R^{k-1}$ } $\to$ we need to construct this.

For every $v \in V_{k-1}$ $\quad g(v) = (f(v), 0)$
and $\quad g(v') = (f(u'), d_{T_k}(v', u'))$

The l_1 distances between any two vertices in V_{k-1} have not changed in this new embedding (since we have simply added a 0 in the new coordinate of all such vertices). Finally, the l_1 distance between any vertex $v \in V_{k-1}$ and v' is

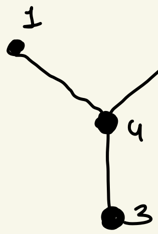
$$d(v, v') = d(v', v') + d(v, v')$$

since there is a unique path between any pair of vertices in a tree.

$\hookrightarrow$ This is isometrically embedded by g.

(If the edge weights are 1, then we can replace d by 1 in the definition of g).

3b (1)



$d(1,4) = d(2,4) = d(3,4) = 1$
$d(i,j) = 2$ otherwise $(i \neq j)$
(shortest path metric)

$(1,2,3,4)$ map to $(f(1), f(2), f(3), f(4))$ on some dimension $\mathbb{R}^n$

Clearly, $\underbrace{\|f(1)-f(2)\|_2 \leq \|f(1)-f(4)\|_2 + \|f(2)-f(4)\|_2}$

For isometric embedding these values are 2, 1, 1

since, there is equality here, it means in the embedded space, they have to be on a single straight line.
(Triangle inequality is strict in $\mathbb{R}^n$ except collinear triples.)
By similar arguments, $\{1,2,4\}$, $\{1,3,4\}$, and $\{2,3,4\}$ are on a single line.
Therefore, all four points must be on the same line which clearly leads to contradiction. Therefore this graph cannot be isometrically embedded into l_2.

(2) We need to show that minimum distortion for embedding this graph into l_2 is $\frac{2}{\sqrt{3}}$. Let Δ denote the triangle formed by $a'b'c'$ where $a' = f(1)$, $b' = f(2)$ and $c' = f(3)$. Next, consider the quantity $\max(\|a'-s'\|, \|b'-s'\|, \|c'-s'\|)$ which lower bounds the distortion of f.
This quantity is maximized when $r = \|a'-s'\| = \|b'-s'\| = \|c'-s'\| \Rightarrow s' =$ center of the smallest enclosing circle of Δ. However, r is minimized when all the edges of Δ are of equal length and are of length $d_G(1,2) = 2$.
Height of the equilateral triangle with side length $2 \Rightarrow h = \sqrt{3}$
radius of inscribing circle $\Rightarrow r = \left(\frac{2}{3}\right)h = \frac{2}{\sqrt{3}}$

$\text{dist}(f) \geq r = \frac{2}{\sqrt{3}}$. This argument is independent of the target dimension d.

4b $X \subseteq U$

X: subset of size k; U: set of size n

The number of possible colours for the first element of X is k.

second elements is $(k-1)$

⋮

last element is 1

Total number of possibilities in which elements of X are coloured with pairwise distinct colors is $(k!)$

overall, number of possibilities $= k^k$ (each element can take one of the k colours)

$$\text{Probability (required)} = \frac{k!}{k^k}$$

By Stirling approximation,

$$\sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{\left(\frac{1}{12n} - \frac{1}{360n^3}\right)} < n!$$

$$\left(\frac{k!}{k^k}\right) > e^{-k} \underbrace{\sqrt{2\pi k}\, e^{\frac{1}{12k} - \frac{1}{360k^3}}}_{>1} > e^{-k}$$

$$\Rightarrow \frac{k!}{k^k} = \text{prob} > e^{-k}$$

Hence, we have proved, the probability that the elements of X are coloured with pairwise distinct colors is at least e^{-k}.

(Derivation of Stirling's Approximation can be found here) $\Rightarrow$ https://en.wikipedia.org/wiki/Stirling%27s_approximation