Homework 5 Solution: Mathematics for AI
For questions, contact TA Seojin Kim (osikjs@kaist.ac.kr), by email.

---

1. The Fundamental Theorem of Markov chains says that for a connected Markov chain, the long-term average distribution $\mathbf{a(t)}= \frac{1}{t}(\mathbf{p(0)} + \mathbf{p(1)} + ... + \mathbf{p(t\text{-}1)})$ converges to a stationary distribution. Does the t step distribution $\mathbf{p(t)}$ also converges for every connected Markov Chain? Prove or disprove it.

solution. t step distribution $\mathbf{p(t)}$ does not necessarily converge.

Counterexample: Consider the transition matrix $P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ and $\mathbf{p(0)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$.

Then, $\mathbf{p(t)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ for even t, and $\mathbf{p(t)} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ for odd t. Therefore $\mathbf{p(t)}$ does not converge.

2. Consider a 2-dimensional lattice with n vertices in each coordinate. The transition probability of each neighbor is $\frac{1}{4}$. Each boundary point has a self-loop with transition probability of $1 - \frac{\text{(number of neighbors)}}{4}$. Find the ε-mixing time for this graph in the following way:

* "Edges leaving S" means the edges s.t. one vertex is in S and the other vertex is not in S.

(a) Show that the minimum number of edges leaving a set S is n when $\frac{n^2}{2} \geq |S| \geq \frac{n^2}{4}$.

(b) Show that the minimum number of edges leaving a set S is $\lfloor 2\sqrt{|S|} \rfloor$ when $|S| \leq \frac{n^2}{4}$.

(c) Compute Φ.

(d) Compute ε-mixing time.

solution.

(a) When $\frac{n^2}{2} \geq |S| \geq \frac{n^2}{4}$, the subset with fewest edges leaving it consists of some number of columns plus one additional partial column. Then, the number of edges leaving S is at least n.

(b) When $|S| \leq \frac{n^2}{4}$, the subset S of a given size that has the minimum number of edges leaving consists of a square located at the lower left hand corner of the grid. If $|S|$ is not a perfect square, then the right most column of S is short. Thus, at least $\lfloor 2\sqrt{|S|} \rfloor$ points in S are adjacent to points in $\bar{S}$.

(c) Note that the stationary distribution $\pi_i = 1/n^2$ for all vertices i. When $\frac{n^2}{2} \geq |S| \geq \frac{n^2}{4}$,

$$\sum_{i \in S} \sum_{j \in \bar{S}} \pi_i p_{ij} \geq \Omega(n \frac{1}{n}) = \Omega(\frac{1}{n}).$$

$$\Phi(S) \geq \Omega(\frac{1/n}{\min(\frac{S}{n^2}, \frac{\bar{S}}{n^2})}) = \Omega(\frac{1}{n}).$$

When $|S| \leq \frac{n^2}{4}$,

$$\sum_{i \in S} \sum_{j \in \bar{S}} \pi_i p_{ij} \geq \frac{\lfloor 2\sqrt{|S|} \rfloor}{n^2}.$$

$$\Phi(S) = \frac{\sum_{i \in S} \sum_{j \in \bar{S}} \pi_i p_{ij}}{\min(\pi(S), \pi(\bar{S}))} \geq \frac{\lfloor 2\sqrt{|S|} \rfloor / n^2}{|S|/n^2} = \Omega(\frac{1}{n}).$$

Therefore, $\Phi = \Omega(\frac{1}{n})$.

(d) By theorem 4.5., the ε-mixing time of a random walk is $O(\frac{\ln(1/\pi_{min})}{\Phi^2 \varepsilon^3})$.

Substituting $\Phi = \Omega(\frac{1}{n})$ and $\pi_{min} = \frac{1}{n^2}$, we have $O(n^2 \ln n / \varepsilon^3)$.

3. Let S be a set of examples, and $\mathcal{H}$ be a hypothesis class. To prove that the VC-dimension of $\mathcal{H}$ is some integer d, we must show: (1) there exists a subset of S of size d shattered by $\mathcal{H}$, and (2) there exist no subsets of S of size $D \geq d$ that are shattered by $\mathcal{H}$.

(a) Prove that if there exists a subset of S of size d shattered by $\mathcal{H}$, then for any $1 \leq k < d$ there also exists a subset of size k of X shattered by $\mathcal{H}$.

* By proving (a), we can relax (2) to the following statement: there exist no subsets S of size $d+1$ that are shattered by $\mathcal{H}$.

(b) Let $S = \mathbb{R}$, and $\mathcal{H}$ be the set of all classifier h that classify a point x as $h(x) = 1$ if $x \in \cup_{i=1}^{p} R_i$, and $h(x) = -1$ otherwise for some set of non-intersecting intervals $R_1, ..., R_p$ with fixed given p. Find the VC-dimension of $\mathcal{H}$.

solution.

(a) Suppose $x_1, ..., x_d \in S$ are shattered by $\mathcal{H}$. Consider the set of $x_1, ..., x_k$ where $1 \leq k < d$. Let $l_1, ..., l_k \in \{-1, 1\}$ be any labeling for $x_1, ..., x_k$. Also, let $l_{k+1} = ... = l_d = 1$. Since $x_1, ..., x_d$ are shattered by $\mathcal{H}$, there exists $h \in \mathcal{H}$ such that $h(x_i) = l_i$ for $i = 1, ..., d$. Therefore, there exists $h \in \mathcal{H}$ such that $h(x_i) = l_i$ for $i = 1, ..., k$. It means, $x_1, ..., x_k$ are shattered by $\mathcal{H}$.

(b) Firstly, we show that $2p$ points can be shattered by $\mathcal{H}$. Consider $x_1 <, ..., < x_{2p}$. Letting $R_i = (x_{2i-1}, x_{2i})$, $R_i = [x_{2i-1}, x_{2i})$, $R_i = (x_{2i-1}, x_{2i}]$, and $R_i = [x_{2i-1}, x_{2i}]$, each pair of points x_{2i-1}, x_{2i} can be shattered with the i'th interval for $i = 1, ..., p$, without affecting the rest. Therefore, $x_1, ..., x_{2p}$ can be shattered.

Secondly, we show that $2p+1$ points can be shattered by $\mathcal{H}$. Consider $x_1 <, ..., < x_{2p+1}$. Label the odd-indexed points 1, and the rest 0. Then, no classifier in $\mathcal{H}$ can be consistent with this labeling.

Combining above results and (a), we have the VC-dimension of $\mathcal{H}$ is $2p$.

4. For $k = 1, 2$, let r_k be some positive integer and $\mathcal{F}_k : \{-1, 1\}^{r_k} \to \{-1, 1\}$ be some function class with VC-dimension d_k. Define $\mathcal{G}_k : \{-1, 1\}^{r_k} \to \{-1, 1\}^{r_{k+1}}$ with $r_3 = 1$ as:

$$\mathcal{G}_k = \{h(x) = (f_1(x), ..., f_{r_{k+1}}(x)) | f_1, ..., f_{r_{k+1}} \in \mathcal{F}_k\}.$$

Our hypothesis class $\mathcal{H}$ is 2-layer feedforward network defined by:

$$\mathcal{H} = \{h_2 \circ h_1 : \{-1, 1\}^{r_1} \to \{-1, 1\} | h_1 \in \mathcal{G}_1, h_2 \in \mathcal{G}_2\}.$$

(a) Prove that the growth function of $\mathcal{H}$ is bounded as:

$$\mathcal{H}[n] \leq \mathcal{F}_1[n]^{r_2} \cdot \mathcal{F}_2[n]$$

(b) Prove $\mathcal{H}[n] \leq (en)^d$, where $d = r_2 d_1 + d_2$.

Hint: Use Sauer's Lemma.

(c) Show that VC-dimension of $\mathcal{H}$ is $O(d \ln d)$.

Hint: Use the result of (b). When $n \geq 16$, we have $\log_2 n \leq \sqrt{n}$.

solution. (a) Firstly, we prove useful lemmas.

Lemma 1. *Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be two function classes with the same domain X. Let $\mathcal{H} = \mathcal{H}_1 \times \mathcal{H}_2$ be their cartesian product. Then,*

$$\mathcal{H}[n] \leq \mathcal{H}_1[n]\mathcal{H}_2[n].$$

Proof. Fix $x^n \in X^n$. Then,

$$|\mathcal{H}(x^n)| = |\mathcal{H}_1(x^n)||\mathcal{H}_2(x^n)| \leq \mathcal{H}_1[n]\mathcal{H}_2[n].$$

As $x^n \in X^n$ was arbitrary, we have $\mathcal{H}[n] \leq \mathcal{H}_1[n]\mathcal{H}_2[n]$. Note that $\mathcal{H}(x^n) := \{(h(x_1), ..., h(x_n)|x^n = (x_1, ..., x_n) \in X^n, h \in \mathcal{H}\}$.

Lemma 2. *Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be two function classes where the range of $\mathcal{H}_1$ and the domain of $\mathcal{H}_2$ coincide. Let $\mathcal{H} = \mathcal{H}_1 \circ \mathcal{H}_2$ be their composition. Then,*

$$\mathcal{H}[n] \leq \mathcal{H}_1[n]\mathcal{H}_2[n].$$

Proof. Let the domain of $\mathcal{H}_1$ and $\mathcal{H}_2$ be X and Y, respectively. Fix $x^n \in X^n$. Then,

$$\begin{aligned}\mathcal{H}(x^n) &= \{(h_2(h_1(x_1)), ..., h_2(h_1(x_n)))|h_1 \in \mathcal{H}_1, h_2 \in \mathcal{H}_2\} \\ &= \cup_{y^n \in \mathcal{H}_1(x^n)} \{(h_2(y_1), ..., h_2(y_n))|h_2 \in \mathcal{H}_2\}.\end{aligned}$$

Therefore,

$$\begin{aligned}|\mathcal{H}(x^n)| &\leq \sum_{y^n \in \mathcal{H}_1(x^n)} |\{(h_2(y_1), ..., h_2(y_n))|h_2 \in \mathcal{H}_2\}| \\ &\leq \sum_{y^n \in \mathcal{H}_1(x^n)} \mathcal{H}_2[n] \\ &= |\mathcal{H}_1(x^n)|\mathcal{H}_2[n] \\ &\leq \mathcal{H}_1[n]\mathcal{H}_2[n].\end{aligned}$$

As $x^n \in X^n$ was arbitrary, we have $\mathcal{H}[n] \leq \mathcal{H}_1[n]\mathcal{H}_2[n]$.

Because $\mathcal{H} = \mathcal{G}_2 \circ \mathcal{G}_1 = \mathcal{F}_2 \circ \mathcal{F}_1^{r_2}$, by lemma 1 and 2, we have $\mathcal{H}[n] \leq \mathcal{F}_1[n]^{r_2} \cdot \mathcal{F}_2$.

(b) By Sauer's lemma, $\mathcal{H}[n] \leq (en/d_1)^{r_2 d_1} \cdot (en/d_2)^{d_2} \leq (en)^{r_2 d_1 + d_2} = (en)^d$.

(c) If the set of size n is shattered by $\mathcal{H}$, then $\mathcal{H}[n] = 2^n$. Note that the largest n with $\mathcal{H}[n] = 2^n$ is the VC-dimension of $\mathcal{H}$.

For $\mathcal{H}[n] = 2^n$, it suffices to have $n \leq d\log_2(en) \leq d\sqrt{en}$ (i.e., $n \leq d^2 e$), for sufficiently large n. By plugging in the original inequality, $n \leq d\log_2(e^2 d^2)$. Hence, the VC-dimension of $\mathcal{H}$ is $O(d\ln d)$.