AI503: Mathematics for Artificial Intelligence

Prof. Jinwoo Shin

Final Exam

(i) Exam is held from 9:00am to 11:45am.
(ii) You should solve the exam questions on a paper (prepared by yourself).
(iii) You should log in to the course zoom link and turn on the camera during the exam to prevent cheating.
(iv) The exam is closed-book and closed-note.
(v) Until 11:45 am, you should scan (or take a picture of) your answers and send the file to jihoontack@kaist.ac.kr. **Late submission is not allowed.**

Problems	Score
Problem 1, (20)	
Problem 2, (20)	
Problem 3, (20)	
Problem 4, (20)	
Problem 5, (20)	
Total (100)	

Problem 1 - (20pt)

Let $X \sim \mathcal{N}(\mu, \sigma^2)$ be a Gaussian random variable (r.v):

$$\text{Prob}(X) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big).$$

(a) [10pt] Prove the following statement: For any $s \in \mathbb{R}$,

$$\mathbb{E}\big[\exp(sX)\big] = \exp\Big(\mu s + \frac{\sigma^2 s^2}{2}\Big).$$

(b) [10pt] Prove the following statement: For any $t > 0$,

$$\text{Prob}(|X-\mu| \geq t) \leq 2\exp\Big(-\frac{t^2}{2\sigma^2}\Big).$$

Problem 2 - (20pt)

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

(a) [7pt] Find the stationary distribution.

(b) [8pt] Compute the normalized conductance Φ.

(c) [5pt] Compute the ε-mixing time.

Problem 3 - (20pt)

For $k = 1, 2, ..., M$, 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 $r_{M+1} = 1$ 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 M-layer feedforward network defined by:

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

(a) [6pt] Prove that the growth function of $\mathcal{H}$ (denoted by $\mathcal{H}[n]$) is bounded as:

$$\mathcal{H}[n] \leq \Pi_{k=1}^{M} (\mathcal{F}_k[n])^{r_{k+1}}$$

where $\mathcal{F}_k[n]$ is the growth function of $\mathcal{F}_k$.

(b) [6pt] Prove $\mathcal{H}[n] \leq (en)^d$, where $d = \sum_{k=1}^{M} r_{k+1} d_k$.

Hint: Use Sauer's Lemma: If d is the VC-dimension of $\mathcal{H}$, then $\mathcal{H}[n] \leq (\frac{en}{d})^d$.

(c) [8pt] Show that VC-dimension of $\mathcal{H}$ is $O(d \ln d)$.

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

Problem 4 - (20pt)

Prove the following statement.

Let $b_1, b_2, \ldots, b_d$ be the distinct values that appear in the input. Select h from the 2-universal family of hash functions. Then the set $S = \{h(b_1), h(b_2), \ldots, h(b_d)\}$ is a set of d random and pairwise independent values from the set $\{0, 1, 2, \ldots, M-1\}$. Show that with probability at least $\frac{2}{3} - \frac{d}{M}$, we have $\frac{d}{6} \leq \frac{M}{\min} \leq 6d$, where min is the smallest element of S.

Problem 5 - (20pt)

Prove the following properties (a, b, c, d) of a graph Laplacian matrix L, which plays an important role in spectral clustering algorithms.

Consder a undirected weighted graph $G = (V, E)$ (with no self-loop), where $V = \{1, 2, \dots, n\}$ is the set of vertices, $E \subset V \times V$ is the set of edges, and $[w_{ij} \geq 0 : (i, j) \in E]$ are weights on edges. The graph Laplacian matrix L is defined as $D - W$, where $D = [D_{ij}]$ is the diagonal matrix such that D_{ii} is the degree of vertex i, defined as $D_{ii} = \sum_{j=1}^{n} w_{ij}$ in the weighted edge case, and W is the weighted adjacency matrix such that $W_{ii} = 0$ and $W_{ij} = \begin{cases} w_{ij} & \text{if } (i,j) \in E \\ 0 & \text{otherwise} \end{cases}$, i.e., the weight of the edge between $i \in V$ and $j \in V$.

(a) [5pt] For every vector $f \in R^n$, we have $f^T L f = \frac{1}{2} \sum_{i,j=1}^{n} w_{ij}(f_i - f_j)^2$.

(b) [5pt] L is symmetric and positive-definite.

(c) [5pt] The smallest eigenvalue of L is 0, and the corresponding eigenvector is the constant $\mathbf{1}$.

(d) [5pt] L has n non-negative, real-valued eigenvalues $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$.