Academic Year of 2018
Admission to the Master's Program
Department of Intelligence Science and Technology
Graduate School of Informatics, Kyoto University
(Fundamentals of Informatics)
(International Course)

13:30 - 15:00, February 7, 2018

NOTES

1. This is the Question Booklet in 3 pages including this front cover.
2. Do not open the booklet until you are instructed to start.
3. After start, check the number of pages and notify proctors (professors) immediately if you find missing pages or unclear printings.
4. This booklet has 2 questions written in English. **Solve all questions.**
5. Write your answers in English, unless specified otherwise.
6. Read carefully the notes on the Answer Sheets as well.

Q. 1 A hash table is an effective data structure for implementing the operations, e.g. INSERT, SEARCH, and DELETE, in computer systems.

1.1 What is the advantage of hash tables compared to directly addressing into an array?
1.2 Given a hash table of size 7 to store integer keys, with linear probing and a hash function $h(x) = x \bmod 7$, show the content of the hash table after inserting the keys 0,11,3,7,1,9 in the given order.
1.3 Given a hash function h to hash n distinct keys into an array T of length m and assuming a uniform hashing, what is the expected cardinality of $\{\{k,l\} : k \neq l \text{ and } h(k) = h(l)\}$?

Q. 2 Breadth first search (BFS) and depth first search (DFS) are two algorithms for traversing trees or graphs.

2.1 Given a set of vertices $\{a,b,c,d,e,s\}$ of a graph, draw the directed graph according to the following vertex adjacency lists:
$adj(s) = [a,c,d], adj(a) = [\,], adj(c) = [b,e], adj(b) = [d], adj(d) = [c], adj(e) = [s]$, where an adjacency list $adj(i)$ denotes the set of neighbors of a vertex i in the graph, and points to the neighbors of i.
2.2 Give the visited vertices in an alphabetical order for the graph given in Q 2.1 using BFS and DFS, respectively. Assume that both algorithms are initially called with the vertex s and that the vertices are stored in the adjacency lists.
2.3 Give a recursive algorithm for DFS in a graph.

An ensemble X is a tuple $(x, \mathcal{A}_X, \mathcal{P}_X)$, where the outcome x is the value of a random variable, which belongs to one of a set of possible values $\mathcal{A}_X = \{a_1, a_2, \ldots, a_i, \ldots, a_\mathcal{I}\}$, having probabilities $\mathcal{P}_X = \{p_1, p_2, \ldots, p_i, \ldots, p_\mathcal{I}\}$, with $P(x = a_i) = p_i \geq 0$ and $\sum_{i=1}^{\mathcal{I}} P(x = a_i) = 1$. $H(X)$ denotes Shannon entropy of ensemble X, and $\mathcal{E}[x]$ denotes the expectation of x.

Q.1 For an arbitrary ensemble X, what is $\mathcal{E}[1/P(x)]$?

Q.2 If f is a convex function, $\mathcal{E}[f(x)] \geq f(\mathcal{E}[x])$ holds. Prove this inequality.

Q.3 Prove $H(X) \leq \log_2(|\mathcal{A}_X|)$ using the inequality in Q.2, where $|\mathcal{A}_X|$ denotes the number of elements in $\mathcal{A}_X$, i.e., $|\mathcal{A}_X| = \mathcal{I}$.

A binary symbol code C for an ensemble X is a mapping from the range of x, $\mathcal{A}_X = \{a_1, \ldots, a_\mathcal{I}\}$ to $\{0,1\}^+$, where $\{\cdot\}^+$ denotes the set of all strings of finite length composed of elements from the set. $c(x)$ denotes the codeword corresponding to x, and $\ell(x)$ denotes its length, with $\ell_i = \ell(a_i)$. The expected length $L(C, X)$ of a binary symbol code C for ensemble X is $L(C, X) = \sum_{x \in \mathcal{A}_X} P(x)\ell(x)$.

Q.4 Let $\mathcal{A}_X = \{\mathsf{a}, \mathsf{b}, \mathsf{c}, \mathsf{d}\}$, $\mathcal{P}_X = \{1/2,\ 1/4,\ 1/8,\ 1/8\}$, and $C = \{\mathtt{0}, \mathtt{01}, \mathtt{011}, \mathtt{111}\}$. Imagine picking one bit at random from the binary encoded sequence $\mathbf{c} = c(x_1)c(x_2)c(x_3)\cdots$. What is the probability that this bit is a $\mathtt{1}$?

Q.5 Let $\mathcal{A}_X = \{\mathsf{a}, \mathsf{b}, \mathsf{c}, \mathsf{d}, \mathsf{e}, \mathsf{f}, \mathsf{g}\}$ and $\mathcal{P}_X = \{0.01, 0.24, 0.05, 0.20, 0.47, 0.01, 0.02\}$. Show a uniquely decodable binary symbol code for the ensemble X such that the expected length $L(C, X)$ is not greater than 2.