2018 Summer Entrance Examination

Department of Creative Informatics
Graduate School of Information Science and Technology
The University of Tokyo

Creative Informatics

INSTRUCTIONS

1. Do not open this brochure until the signal to begin is given.
2. Write your examinee ID number below on this cover page.
3. Answer all three problems.
4. Three answer sheets are given. Use a separate sheet of paper for each problem. You may write on the back of the sheet.
5. Write your examinee ID number and the problem number inside the top blanks of each sheet.
6. Do not bring the answer sheets or this brochure out of this room.

Examinee ID ____________________________

This page is blank.

This page is blank.

Problem 1

A data set S including eight data is given as in Figure 1. Each datum is in the form of $(x_1, x_2, x_3, x_4, y) \in \{0,1\}^5$. Below we consider how to construct a rule from S for classifying $\mathbf{x} = (x_1, x_2, x_3, x_4)$ into $y = 1$ or $y = 0$. Answer the following questions.

t	x_1	x_2	x_3	x_4	y
1	1	0	0	1	1
2	1	1	0	0	1
3	1	0	1	1	1
4	0	1	1	0	1
5	0	0	1	1	0
6	0	0	1	0	0
7	0	0	0	1	0
8	1	1	1	0	0

Figure 1: Data set

(1) Let us consider the code-length required for encoding a binary string $\mathbf{z} = z_1, \ldots, z_n$ of length n ($z_i \in \{0,1\}$, $i = 1, \ldots, n$). In general, for a finite set G, the code-length required for encoding one element $g \in G$ is given by $\log |G|$ (bit), where the logarithm is to the base 2. Show that, for a binary string $\mathbf{z}$ of length n in which the number of occurences of 1 is k, the code-length required for encoding $\mathbf{z}$ and k itself is at most

$$\log(n+1) + \log {}_n\mathrm{C}_k \quad \text{(bit)}, \tag{1}$$

where the value of the number n is given in advance and the code-length can be non-integer valued.

Below the value of Eq.(1) for $\mathbf{z}$ is denoted as $L(\mathbf{z})$.

(2) Let the probability of $y = 1$ for a datum with $x_1 = 1$ be θ. Then calculate the least squares estimate of θ from Figure 1. The least squares estimate of θ is the value of θ that minimizes $\sum_{t:x_1=1}(y(t) - \theta)^2$ where $y(t)$ denotes the value of y for the t-th datum and $\sum_{t:x_1=1}$ denotes the sum taken over all the data such that $x_1 = 1$.

(3) Let $\mathbf{y}$ be a binary string obtained by concatenating the values of y for all the data in S in Figure 1. Let $\mathbf{y}_1^{(i)}$ be a binary string obtained by concatenating the values of y for all the data in S such that $x_i = 1$ and let $\mathbf{y}_0^{(i)}$ be a binary string obtained by concatenating the values of y for all the data in S such that $x_i = 0$ ($i = 1, 2, 3, 4$). For example, the indexes of data in S such that $x_1 = 1$ are $1, 2, 3, 8$, so the binary string obtained by concatenating the values of y corresponding to them is $\mathbf{y}_1^{(1)} = 1110$. We define the measure of goodness of classifying data by partitioning S based on whether $x_i = 1$ or $x_i = 0$ as follows:

$$\Delta(i|\mathbf{y}) \stackrel{\text{def}}{=} L(\mathbf{y}_1^{(i)}) + L(\mathbf{y}_0^{(i)}) \quad (i = 1, 2, 3, 4). \tag{2}$$

We consider that the smaller the value of Eq.(2) is, the more the value of x_i contributes to the classification of y. Find i that minimizes $\Delta(i|\mathbf{y})$. Hereinafter, when there are more than one "i"s that minimize the value of Eq.(2), one is chosen randomly from among them.

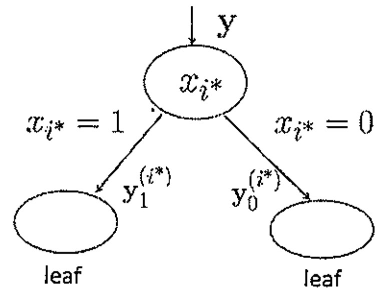

Figure 2: Partitioning tree

(4) Let the value of i obtained in Question (3) be i^*. S is partitioned into two sets according to whether $x_{i^*} = 1$ or $x_{i^*} = 0$, then y is also partitioned into two strings: $\mathrm{y}_1^{(i^*)}$ and $\mathrm{y}_0^{(i^*)}$. It can be represented using a tree structure as shown in Figure 2. We call it a partitioning tree. We call $\mathrm{y}_1^{(i^*)}$ and $\mathrm{y}_0^{(i^*)}$ partitioned strings. We further partition each of $\mathrm{y}_1^{(i^*)}$ and $\mathrm{y}_0^{(i^*)}$, by finding $i(\neq i^*)$ minimizing $\Delta(i|\mathrm{y}_1^{(i^*)})$ and minimizing $\Delta(i|\mathrm{y}_0^{(i^*)})$, respectively. Let this partitioning of a leaf be repeated until the following stopping rule is fulfilled: The depth of a leaf (the number of partitionings from the root to the leaf) is two, or the partitioned string arriving at a leaf is all $y = 1$ or all $y = 0$. Find the partitioning tree that is finally obtained.

(5) For the resulting partitioning tree, for a partitioned string arriving at each leaf, we assign $y = 1$ to the leaf if the number of occurences of $y = 1$ in this string is larger than that of $y = 0$, and assign $y = 0$ to the leaf if the number of occurrences of $y = 1$ is smaller than that of $y = 0$. When the number of occurences of $y = 1$ is the same as that of $y = 0$, we assign randomly $y = 1$ or $y = 0$ to the leaf. This tree can be used for predicting the value of y for any new datum. That is, when (x_1, x_2, x_3, x_4) in the new datum is given and arrives at a leaf, the tree predicts the value of its corresponding y as the value of y assigned to the leaf. Here, even if we change the stopping rule in Question (4) to construct a larger tree from a training data set S so that the values of y for data reaching at each leaf are all $y = 1$ or all $y = 0$, such a tree doesn't necessarily predict the value of y for a new datum with higher accuracy. Explain the reason.

(6) Consider a general case where for a positive integer $d \geq 2$, a set S of multi dimensional data in the form of $(x_1, \ldots, x_d, y) \in \{0,1\}^{d+1}$ and a partioning tree $\mathcal{T}$ are given. Let $\mathcal{M}$ be a set of all subtrees which share the root of $\mathcal{T}$ and are obtained by pruning $\mathcal{T}$ starting from its leaves. We define the following penalized criterion for evaluating the goodness of a subtree $M \in \mathcal{M}$ for the given S:

$$N_L(M)C_L + N_I(M)C_I + \sum_u L(\mathrm{y}_u), \tag{3}$$

where $N_L(M)$ is the total number of leaves in M and $N_I(M)$ is the total number of inner nodes in M. C_L and C_I are given positive constants. The sum in the third term in Eq.(3) is taken over all the leaves $\{u\}$ in M and y_u is the binary string obtained by concatenating the values of y for all the data which reach the leaf u. The smaller the value of Eq.(3) is, the better M is. Give an algorithm that finds M minimizing the criterion Eq.(3) from $\mathcal{M}$ and S, and runs as efficiently as possible in computation time.

Problem 2

Let us consider to control the position of a cart with mass M placed on a slope with angle θ as illustrated in Figure 1. We can move the cart by force f along the x-axis parallel to the slope. Assume that f can be sufficiently large to pull up the cart. Friction between the cart and the slope and air resistance are negligible. We let $f(t)$, $x(t)$, and $v(t)$ denote the force f, the position, and the velocity of the cart at time t, respectively. The magnitude of gravity acceleration is denoted by g.

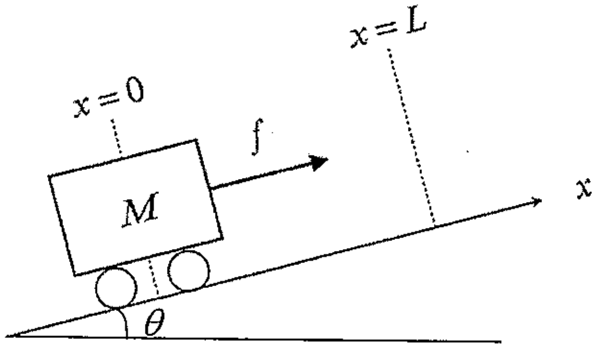

Figure 1.

Suppose that $x(0)=0$ and $v(0)=0$ at time $t=0$. We consider a method to move the cart to the position $x=L$.

Answer the following questions.

(1) Find the position and the velocity of the cart when we accelerate it with a constant force $f(t)=F\ (F>0)$ until the time t_1.

(2) We want to deaccelerate the cart with a constant force $f(t)=-F\ (F>0)$ from the time t_1 in Question (1) until the time $t_2\ (t_2 \geq t_1)$ so that $x(t_2)=L$ and $v(t_2)=0$. Find t_1 and t_2 which realize this motion.

Next, we consider to give a force proportional to the displacement from the target position $x=L$. Specifically, we give $f(t)=k_1\{L-x(t)\}$. k_1 is a positive constant.

(3) Write down the equations of motion for this case.

(4) Draw a graph of $x(t)$.

Next, we consider to further add a force proportional to the velocity of the cart. Specifically, we give $f(t)=k_1\{L-x(t)\}-k_2 v(t)$. k_1 and k_2 are positive constants.

(5) Explain an effect caused by adding $-k_2 v(t)$ and the reason why this effect occurs.

(6) Find the condition regarding k_1 and k_2 so that $x(t)$ does not oscillate. You can use the following facts if necessary.

> The general solution of a differential equation
>
> $$\frac{d^2x}{dt^2}+a\frac{dx}{dt}+bx=0 \quad (a \text{ and } b \text{ are real-valued constants})\cdots(\mathrm{A})$$
>
> can be represented by the solution of the quadratic equation
>
> $r^2+ar+b=0\cdots(\mathrm{B})$
>
> as follows:
>
> 1. When Eq.(B) has two different real roots p and q,
> $x=C_1e^{pt}+C_2e^{qt}$.
> 2. When Eq.(B) has two different imaginary roots $h\pm ki$,
> $x=e^{ht}(C_1\cos kt+C_2\sin kt)$.
> 3. When Eq.(B) has a double root p,
> $x=e^{pt}(C_1+C_2t)$.
>
> Here, C_1 and C_2 are constants of integration.

(7) Draw a graph of $x(t)$ under the condition obtained in Question (6).

Next, we consider to further add a force proportional to the integral of the displacement from the target position. Specifically, we give $f(t)=k_1\{L-x(t)\}-k_2v(t)+k_3\int_0^t\{L-x(\tau)\}d\tau$. k_1, k_2, and k_3 are positive constants.

(8) Explain an effect caused by adding $k_3\int_0^t\{L-x(\tau)\}d\tau$ and the reason why this effect occurs.

Problem 3

Select four items out of the following eight items concerning information systems, and explain each item in approximately from four to eight lines of text. If necessary, use examples or figures.

(1) Pipeline hazard

(2) Register renaming

(3) Kalman filter

(4) Regular grammar and regular languages

(5) Public-key cryptography and certification authority

(6) Traveling salesman problem

(7) Divide and conquer method

(8) Vector quantization

This page is blank.

This page is blank.