2014 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 <u>three problems out of the four</u>.
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 down the examinee ID number and the problem number inside the top blanks of each sheet.
6. Do not take out the answer sheets or this brochure from this room.

Examinee ID ______________________________

This page is blank.

This page is blank.

Question 1

We consider a weather prediction system in which a single senior predictor predicts a weather probability distribution on the basis of N predictors' prediction results. Below the system is described in details.

There are N weather predictors, each of whom outputs a weather probability distribution once a day. Here the weather is a binary random variable taking a value 1 or 0 only (1 means "fine" while 0 means "not fine"). It is assumed that the weather is independent of a day.

Let the probability distribution that the i-th predictor outputs on the t-th day be $P_i^{(t)}(X)$ $(X \in \{1,0\})$ where we let $0 < P_i^{(t)}(X) < 1$ $(X \in \{1,0\})$. There is a senior predictor who aggregates the outputs of the N predictors. On the t-th day, the senior predictor takes a weighted average over the probability distributions output by the N predictors to output a weather probability distribution $\hat{P}^{(t)}(X)$ $(X \in \{1,0\})$. Here the weight on the i-th predictor on the t-th day is denoted as $v_i^{(t)}$ $(\sum_{i=1}^{N} v_i^{(t)} = 1, v_i(t) > 0\ (i = 1, \ldots, N))$. That is, on the t-th day, $\hat{P}^{(t)}(X)$ is given by $\sum_{i=1}^{N} v_i^{(t)} P_i^{(t)}(X)$ (see Figure 1).

On the t-th day, after the senior predictor ouputs, the real outcome $x_t \in \{1,0\}$ of the weather on the day is reported. This process goes on sequentially with respect to t.

In the above setting, answer the following questions.

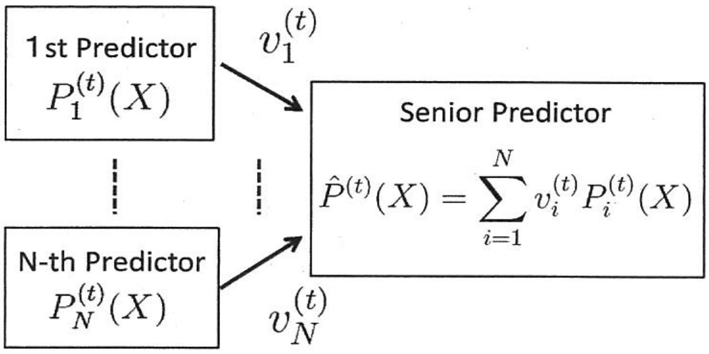


Figure 1: Weighted Average Prediction of Senior Predictor

(1) Assume that the senior predictor defines the weight $v_i^{(t)}$ of the i-th predictor on the t-th day so that the ratio of $v_i^{(t)}$s with respect to i is equal to that of their corresponding likelihoods with respect to the sequence of the past $t-1$ data: $x^{t-1} = x_1 \ldots x_{t-1}$. That is, when we denote the likelihood of the i-th predictor with respect to x^{t-1} as $w_i^{(t-1)}$, the following equation holds: For each t,

$$v_1^{(t)} : \cdots : v_N^{(t)} = w_1^{(t-1)} : \cdots : w_N^{(t-1)}.$$

The likelihood of the i-th predictor with respect to x^{t-1} is calculated as $w_i^{(t-1)} = \prod_{j=1}^{t-1} P_i^{(j)}(X = x_j)$, where we denote $P_i^{(j)}(X = x_j)$ as $P_i^{(j)}(x_j)$ and we set $P_i^{(0)}(x_0) = 1$ $(i = 1, \ldots, N)$.

In this setting, for each i, show a relation between $w_i^{(t)}$ and $w_i^{(t-1)}$, and derive a formula for calculating $v_i^{(t+1)}$ using $w_i^{(t-1)}$ and $P_i^{(t)}(x_t)$ $(i = 1, \ldots, N)$.

(2) Suppose that the senior predictor predicts the weather probability distribution sequentially with respect to t for T days. Then under the setting of $v_i^{(t)}$ as in (1), write an algorithm for the senior predictor to output a weather probability distribution and update the weights for predictors every day, and show the order of its computation time in terms of N and T. Here the initial weight for each predictor is set as follows:

$$v_1^{(0)} = \ldots = v_N^{(0)} = 1/N.$$

(3) When the prediction is sequentially made for T days, we define the cumulative predictive loss for the senior predictor with respect to the sequence $x^T = x_1 \ldots x_T$ of observed real coutcome as follows:

$$Loss(x^T) = \sum_{t=1}^{T} \left(-\log \hat{P}^{(t)}(x_t) \right).$$

Then write $Loss(x^T)$ as a function of $P_i^{(t)}(x_t)$ $(i = 1, \ldots, N,\ t = 1, \ldots, T)$ and N. Here the logarithm is the natural logarithm.

(4) Prove that the senior predictor's cumulative loss for T days defined in (3) is at most $\log N$ larger than the least cumulative loss over all i-th predictors for T days. Here the cumulative loss for the i-th predictor for T days is defined as $\sum_{t=1}^{T}(-\log P_i^{(t)}(x_t))$ $(i = 1, \ldots, N)$.

Question 2.

Answer the following questions on virtual memory by paging.

The system consists of a processor, a main memory and a secondary storage (e.g. magnetic disks). The processor issues memory accesses using logical addresses. When the target of access is located in the main memory, the logical address is translated into the physical address by an address translation table. When the target of access is not located in the main memory, a page in the main memory is selected, and the selected page is swapped with the page in the secondary storage that contains the target of the access. The new physical address is then added to the address translation table.

Following is a glossary:

- Page miss ratio: Ratio of memory accesses where the targets are not in the memory among all the memory accesses.
- Page replacement algorithm: The algorithm that selects a page to be swapped into the secondary storage to assign a new page on the main memory.
- LRU (Least Recently Used): One of page replacement algorithms. A page that was accessed at the oldest time is selected for replacement.
- FIFO (First In First Out): One of page replacement algorithms. A page that was assigned to the main memory at the oldest time is selected for replacement.

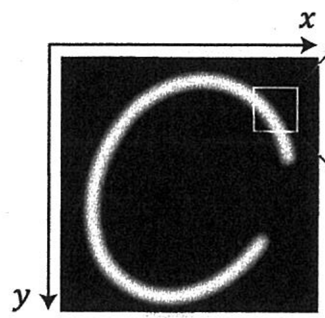
(1) Assume that the access time of main memory is Tm and the access time of secondary storage is Ts. The size of page is 4KB, and no time is taken for all data transfer between the main memory and the secondary storage. Answer the page miss ratio and the average memory access time when the processor sequentially accesses the vector data. Here, the size of the vector is much larger than the size of main memory. Then, calculate the actual values when Tm = 100ns and Ts = 1ms.

(2) Ordinary programs repeatedly access same memory data. In this case, the page replacement algorithm has large influence on the performance. LRU (Least Recently Used) and FIFO (First In First Out) are typical page replacement algorithms. Show an example of program structure where the LRU page replacement algorithm gives better performance than the FIFO algorithm.

(3) It is difficult to implement exact LRU algorithm for page replacement. For this reason, an approximation algorithm is used instead. Show an approximation algorithm of LRU for page replacement.

Question 3

We consider the problem of distinguishing between two hand-written alphabets "C" and "I" using image processing and pattern recognition techniques. Figure 1 shows an input image X. Let i and j denote integers corresponding to the x and y coordinates of a pixel. The value of a pixel (i,j) is represented by $f(i,j)$. Here, we take the top-left pixel as the origin $(0,0)$ of the coordinates.



200	30	30	0	0
200	200	80	30	0
80	250	$\boxed{150}$	80	30
30	200	250	200	30
30	30	200	200	200

Figure 1： Input image X (Numbers represent pixel values)

Following the procedure below, we extract some features to quantify the properties of the input image. First, we apply a 3×3 size linear filter. Let $g(i,j)$ denote the output value of the filter for a pixel (i,j), which is defined as

$$g(i,j) = \sum_{n=-1}^{1} \sum_{m=-1}^{1} f(i+m, j+n)h(m,n),$$

where $h(m,n)$ is the array that contains the coefficients of the filter (m and n are integer values). Answer the following questions.

(1) We apply the linear filter L illustrated in Figure 2 to the image X. Compute the output value for the pixel P in Figure 1.

$h(-1,-1) = 0$	$h(0,-1) = 1$	$h(1,-1) = 0$
$h(-1,0) = 1$	$h(0,0) = -4$	$h(1,0) = 1$
$h(-1,1) = 0$	$h(0,1) = 1$	$h(1,1) = 0$

Figure 2： Linear filter L

Next, we reduce the resolution of the filtered image to 6×6 pixels and binarize its values. Specifically, the converted image is represented as $f(i,j) = \{0,1\}$ $(0 \le i,j \le 5)$. Figure 3 shows two examples of binary images Y1 and Y2. We extract features called "image moments" from them. The image moment of order $(p+q)$ is defined as $M_{pq} = \sum_{i,j} i^p j^q f(i,j)$. For example, we can see that M_{00} corresponds to the $\boxed{\text{(A)}}$ of the figure in the binary image, and $(M_{10}/M_{00}, M_{01}/M_{00})$ corresponds to its $\boxed{\text{(B)}}$.

(2) Fill in the blanks (A) and (B) with appropriate words. Also, compute the image moments M_{00} and M_{10} of Y1 and Y2 respectively.

0	0	1	1	1	0
0	1	0	0	0	1
1	1	0	0	0	0
1	0	0	0	0	0
0	1	0	0	1	0
0	0	1	1	0	0

Y1

0	0	1	1	1	0
0	0	0	1	0	0
0	0	0	1	0	0
0	0	1	1	0	0
0	0	1	0	0	0
0	1	1	1	0	0

Y2

Figure 3 : Binary images Y1 (Left) Y2 (Right)

Taking a pair of two features (M_{00}, M_{10}) extracted by the above procedure as an input pattern, we classify this into one of two classes ("C" or "I"). Table 1 lists the values of features of four "C" and four "I" training examples. Also, Figure 4 shows their plots on a two-dimensional graph.

Table 1 : Training examples

	C1	C2	C3	C4	I1	I2	I3	I4
M_{00}	10	12	12	14	8	10	10	12
M_{10}	24	24	26	26	27	26	28	27

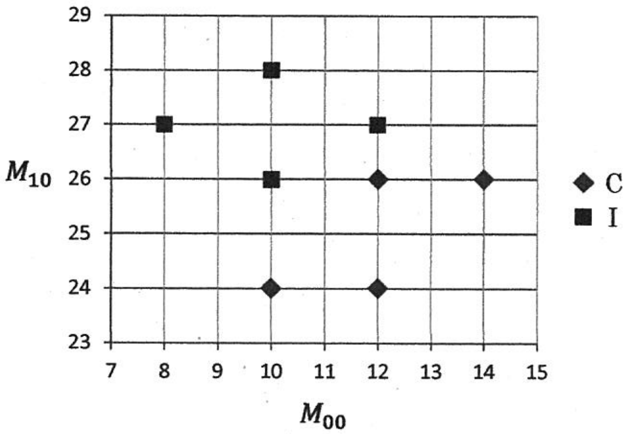


Figure 4 : Training examples (Plots)

(3) We classify an input pattern using the k nearest neighbor (k-NN) method. Namely, a pattern is classified into the class that has the largest number of samples among its k nearest training examples. Suppose we have a new pattern of an input image with the features $(M_{00}, M_{10}) = (13, 27)$. Classify this pattern into "C" or "I" using the k-NN method and describe the reason when $k = 1$ and $k = 3$, respectively. Distance between any two patterns is defined in terms of the Euclidean distance.

(4) As an alternative classification algorithm, we classify an input pattern into the class where the mean of its training examples is nearest to the pattern in terms of the Euclidean distance. Show the equation that defines the boundary of discrimination, and the methodology of discrimination using it. Also, classify the pattern $(M_{00}, M_{10}) = (13, 27)$ using the methodology.

(5) Compare the two algorithms described at (3) and (4) and discuss their advantages respectively.

Question 4

Select <u>four items</u> out of the following eight items concerning information systems, and explain each item in approximately 4~8 lines of text. If necessary, use examples or figures.

(1) tf-idf
(2) ZMP (Zero Moment Point)
(3) Distributed hash
(4) Shortest path problem
(5) Bayesian network
(6) Carry look ahead
(7) Closure
(8) Finite automaton

This page is blank.

This page is blank.