2024年度夏入試 / 2024 Summer Entrance Examination

東京大学情報理工学系研究科

創造情報学専攻

創造情報学

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

Creative Informatics

注意事項

1. 試験開始の合図まで, この問題冊子を開かないこと.
2. この表紙の下部にある受験番号欄に受験番号を記入すること.
3. 3問全てに, 日本語ないし英語で回答すること.
4. 解答用紙は3枚配られる. 1問ごとに必ず1枚の解答用紙を使用すること. 解答用紙のおもて面に書ききれないときには, うら面にわたってもよい.
5. 解答用紙の指定された箇所に, 受験番号および問題番号を忘れずに記入すること.
6. 解答用紙および問題冊子は持ち帰らないこと.

INSTRUCTIONS

1. Do not open this booklet until the start of the examination is announced.
2. Write your examinee's number below on this cover page.
3. Answer all three problems in Japanese or English.
4. Three answer sheets are given. Use a separate answer sheet for each problem. You may write on the back of the answer sheet.
5. Write your examinee's number and the problem number inside the top blanks of each sheet.
6. Do not bring the answer sheet or this booklet out of this room.

受験番号 / Examinee's number ______________________________

このページは空白.

This page is blank.

このページは空白.

This page is blank.

第１問

以下ではある確率過程に関して、N個の観測値列からなる標本$\mathrm{X}=(x_1,x_2,x_3,\ldots,x_N)$を得た時、そこから元の確率情報源のパラメータを最尤推定によって求めることを考える。最尤推定とはパラメータ推定法の一種であり、以下で説明される。

パラメータθに従う分布の密度関数を$f(x;\theta)$とする。標本Xを得た時の尤度関数$L(\mathrm{X}\|\theta)$は$\prod_{n=1}^{N} f(x_n;\theta)$で与えられる。ここで、$L(\mathrm{X}\|\theta)$を最大にするような$\theta$を最尤推定量という。

以下の問いに答えよ。

(1) 一つの例として、コイントスを想定する。ある１枚のコインの表が出る確率をθとする。このコインをN回投げた時、r回表が出るという観測を得た。この時の尤度関数を上記の定義に従って求めよ。

(2) 最尤推定によって表が出る確率θを推定する。まず尤度関数の対数値（これを以後、対数尤度関数と呼ぶ）を考え、次に微分によってその対数尤度関数を最大にするθを求めよ。

平均μ、分散σ^2というパラメータを持つ正規分布を$\mathcal{N}(\mu,\sigma^2)$と記述し、その確率密度関数は以下で与えられるものとする。

$$f(x;\mu,\sigma^2)=\frac{1}{\sqrt{2\pi}\sigma}e^{-(x-\mu)^2/(2\sigma^2)}$$

複数の正規分布から生成されるデータを観測した場合を考える。これは Gaussian Mixture Model (GMM) と呼ばれる。GMM のデータ生成過程は以下のように記述される。

１．番号k $(k=1,\ldots,K)$を確率π_kで選択する分布（これをカテゴリカル分布 C と呼ぶ）を想定する。またこの番号に対応して、K個の正規分布 $\mathcal{N}_1(\mu_1,\sigma_1^2)$, …, $\mathcal{N}_K(\mu_K,\sigma_K^2)$を想定する。

２．標本Xの要素$(n=1,\ldots,N)$ごとに

　２－I　上で述べたカテゴリカル分布 C に従って番号z_nを生成

　２－II　生成された番号z_nに対応する正規分布 $\mathcal{N}_{z_n}(\mu_{z_n},\sigma_{z_n}^2)$に従って$x_n$を生成

３．生成されたN個の観測値列$(x_1,x_2,x_3,\ldots,x_N)$が標本となる。

この標本から GMM のパラメータ $\Theta=\{\mu_k,\sigma_k,\pi_k\}_{k=1,\ldots,K}$を最尤推定で求める。

まず$K=1$の場合（単一の正規分布からのデータ生成）を考える。この正規分布から生成されたN個の観測値列からなる標本$\mathrm{X}=(x_1,x_2,x_3,\ldots,x_N)$を得た時、パラメータ（平均$\mu$、分散$\sigma^2$）を最尤推定したい。以下の問いに答えよ。

(3) この場合の対数尤度関数を求めよ。

(4) 平均μに関して対数尤度関数を最大化する値（平均μの最尤推定量）を求めよ。また同様にして、分散σ^2に関して対数尤度関数を最大化する値（分散σ^2の最尤推定量）を求

めよ。

次に$K \geq 2$の場合を考える。以下の問いに答えよ。

(5) このときの対数尤度関数をパラメータ Θ と $Z = (z_1, z_2, \ldots, z_N)$の関数として記述したい。まず、Zが生成される確率 $p(Z|\Theta)$をπ_kを使って記述せよ。次に、Zが与えられたもとでのXの生成確率 $p(X|Z,\Theta)$を記述せよ。最後に、前記で求めた確率 $p(Z|\Theta)$および $p(X|Z,\Theta)$を使って対数尤度関数を記述せよ。

前記の問い(5)で求めた対数尤度関数の最大化を行いたい。しかし、この対数尤度関数においては対数関数の内に和演算Σが入っているので、導関数が零となる Θ を求めることは困難である。そこで以下の「アルゴリズム A」を用いて反復的に最大化を行う（なおここでは局所的最大化に留め、大域的最大化までは求めない）。アルゴリズム A においては、最大化したい目的関数 $D(\Theta)$に関して $D(\Theta) = \max_\theta G(\Theta, \theta)$となる補助関数 $G(\Theta, \theta)$および補助変数 θ を用いて、[step 1] $\theta \leftarrow \mathrm{argmax}_{\theta'} G(\Theta, \theta')$, [step 2] $\Theta \leftarrow \mathrm{argmax}_{\Theta'} G(\Theta', \theta)$、という操作を反復的に繰り返して Θ の局所最適解を計算する。以下の問いに答えよ。

(6) 上記のアルゴリズム A を用いる場合、step 1～step 2 の繰り返しで更新される Θ は必ず元の目的関数 $D(\Theta)$を上昇もしくは停留させることを示せ。

(7) 前記の問い(5)で求めた対数尤度関数の補助関数を求めるために、Jensen の不等式を利用する。対数関数に関する Jensen の不等式は $\log(\Sigma_i \lambda_i y_i) \geq \Sigma_i \lambda_i \log(y_i)$ （ここでy_iは正の変数、λ_iは$\Sigma_i \lambda_i = 1$, $\lambda_i \geq 0$ を満たす任意の重み）と書ける。この不等式を証明せよ。

(8) Jensen の不等式を用いて、問い(5)で求めた対数尤度関数の補助関数を設計せよ。ここではλ_iを補助変数とせよ。和演算Σの中にある個別の正規分布を一旦λ_iで割り、かつその前にλ_iを乗ずることによって補助関数を導出すると良い。

(9) 前記の問い(8)で求めた補助関数に関し、アルゴリズム A を実行する。まず、[step 1]として補助関数を最大化するλ_iを求めよ。次に[step 2]として（補助変数は固定し）補助関数を最大化するパラメータ Θ を求めよ。

第2問

長さb ($b \geq 1$)のビット列$\boldsymbol{x} = (x_1, x_2, \ldots, x_b)$ ($x_i \in \{0,1\}, 1 \leq i \leq b$) の形式で表されるデータ$N$個が構成するデータセットを考える。各データには、重複しない整数による固有のデータID（識別子）が振られている。任意の入力データ（クエリデータ）と近い距離にあるデータを検索するシステムを作りたい。検索では、条件を満たす全てのデータのデータ ID を列挙する必要がある。二つのデータ間の距離は、ビット列の間のハミング距離によって定義されるものとする。ここで、二つのビット列$\boldsymbol{x} = (x_1, x_2, \ldots, x_b)$と$\boldsymbol{y} = (y_1, y_2, \ldots, y_b)$の間のハミング距離は以下のように定義される。

$$\mathrm{d}(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{b} |x_i - y_i|$$

以下の問いに答えよ。

(1) 以下の表に、$b = 4$、$N = 3$の場合のデータセットの例を示す。

データ ID	x_1	x_2	x_3	x_4
1	0	1	1	1
2	1	0	0	1
3	0	0	1	0

今、クエリデータ$(1,1,0,1)$が与えられたとする。クエリデータと各データの間のハミング距離を求めよ。

次に、ルックアップテーブルを用いた検索アルゴリズムを考える。ただし、bは偶数であり、データのビット列は一様に分布しているものとする。また、以下の問いではルックアップテーブルの構築にかかる時間計算量は考慮しなくてよい。

(2) 与えられたクエリデータと同じビット列を持つデータを検索したい。この時、ビット列を入力とし、そのビット列を持つ全てのデータのデータ ID を納めたリストを出力するルックアップテーブルを考える。このルックアップテーブルを用いた検索の平均時間計算量と空間計算量を答えよ。

(3) ビット列を長さ$b/2$の二つのビット列に分割し、分割されたビット列のそれぞれについて問い(2)と同様にルックアップテーブルにより候補を検索したのちに、クエリデータと一致する全てのデータのデータ ID を返すアルゴリズムを考える。このアルゴリズムによる検索の平均時間計算量と空間計算量を答えよ。

(4) 問い(3)のデータ構造を用い、与えられたクエリデータとのハミング距離が 1 以下のデータを検索するアルゴリズムを考える。このアルゴリズムによる検索の平均時間計算量を答えよ。

次に、$2^b \gg N$の場合を考える。この時、クエリデータと各データの間のハミング距離を実際に計算し、線形探索を行うことが有効となる場合がある。このため、ハミング距離を計算する専用のデジタル回路を設計することを考える。

(5) 二つの 1 ビット列 $\boldsymbol{x} = (x_1)$、$\boldsymbol{y} = (y_1)$ ($x_1, y_1 \in \{0,1\}$) を入力とし、$\boldsymbol{x}$と$\boldsymbol{y}$の間のハミング距離$z \in \{0,1\}$を出力する 2 入力 1 出力のデジタル回路H_1を考える。$\boldsymbol{x}, \boldsymbol{y}, z$の関係を表で表せ。また、$H_1$を AND、OR、NOT ゲートのうち必要なものを用いて描け。

(6) 二つの 2 ビット列間のハミング距離の 2 進表現を出力する 4 入力 2 出力のデジタル回路H_2を AND、OR、NOT ゲートおよびH_1のうち必要なものを用いて描け。

(7) 二つの 4 ビット列間のハミング距離の 2 進表現を出力する 8 入力 3 出力のデジタル回路H_4をH_2、半加算器 HA および OR ゲートのうち必要なものを用いて描け。

第３問

以下に示す情報システムに関する８項目から４項目を選択し、各項目を 4～８行程度で 説明せよ。 必要に応じて例や図、数式を用いてよい。

(1) 動的計画法
(2) ゼロモーメントポイント(ZMP)
(3) BNF 記法(バッカス・ナウア記法あるいはバッカス・ノーマル記法)
(4) 広域ネットワークにおけるトランスペアレント キャッシュ
(5) 自動運転車システムにおけるダイナミック マップ
(6) スレッドレベル並列投機実行
(7) プロシージャル モデリング
(8) k 近傍法

このページは空白.

This page is blank.

Problem 1

In the following, when a sample $X = (x_1, x_2, x_3, ..., x_N)$ consisting of N observations is obtained for a stochastic process, we calculate the parameters of the original stochastic information source by using maximum likelihood estimation. Maximum likelihood estimation is a type of parameter estimation method and is described below.

Let $f(x;\theta)$ be the density function of the distribution in accordance with the parameter θ. The likelihood function $L(X\|\theta)$ when obtaining a sample X is given by $\prod_{n=1}^{N} f(x_n;\theta)$. Here, θ that maximizes $L(X\|\theta)$ is called a maximum likelihood estimator.

Answer the following questions.

(1) As an example, consider the coin toss. Let θ be the probability of a coin coming up faces. We obtained the observation that when this coin is tossed N times, faces come up r times. Obtain the likelihood function in this case based on the above definition.

(2) Estimate the probability θ of a coin coming up faces by using maximum likelihood estimation. First, consider the logarithmic value of the likelihood function (hereafter referred to as the *log-likelihood function*), and then find θ that maximizes the log-likelihood function by differentiation.

A normal distribution with parameters, the mean μ and the variance σ^2, is described as $\mathcal{N}(\mu, \sigma^2)$, and its probability density function is given by

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}.$$

Consider the case of observing data generated from multiple normal distributions. This is called the *Gaussian Mixture Model (GMM)*. The data generation process of the GMM is described as follows.

1. Assume a distribution that selects the number k $(k = 1, ..., K)$ with probability π_k (this is called a categorical distribution C). Also, assume K normal distributions $\mathcal{N}_1(\mu_1, \sigma_1^2)$, ..., $\mathcal{N}_K(\mu_K, \sigma_K^2)$, each of which corresponds to each number.
2. For each element $(n = 1, ..., N)$ of sample X;
   - 2-I: generate number z_n in accordance with the categorical distribution C mentioned above,
   - 2-II: generate x_n in accordance with the normal distribution $\mathcal{N}_{z_n}(\mu_{z_n}, \sigma_{z_n}^2)$ corresponding to the generated number z_n.
3. The generated sequence of N observation values $(x_1, x_2, x_3, ..., x_N)$ is a sample.

From this sample, we obtain the GMM parameters $\Theta = \{\mu_k, \sigma_k, \pi_k\}_{k=1,...,K}$ by maximum likelihood estimation.

First, consider the case of $K = 1$ (data generation from a single normal distribution). When we obtain a sample $X = (x_1, x_2, x_3, ..., x_N)$ consisting of N observations generated from this normal distribution, we want to estimate the parameters (mean μ and variance σ^2) by maximum likelihood estimation. Answer the following questions.

(3) Find the log-likelihood function for this case.

(4) Find the value that maximizes the log-likelihood function with respect to the mean μ (i.e., find the maximum likelihood estimator of the mean μ). Similarly, find the value that maximizes the log-likelihood function with respect to the variance σ^2 (i.e., find the maximum likelihood estimator of the variance σ^2).

Next, consider the case of $K \geq 2$. Answer the following questions.

(5) We would like to describe the log-likelihood function in this case as a function of the parameters Θ and $Z = (z_1, z_2, \ldots, z_N)$. First, describe the probability $p(Z|\Theta)$ that Z is generated using π_k. Next, given Z, describe the generation probability $p(X|Z,\Theta)$ of X. Finally, describe the log-likelihood function using the probabilities $p(Z|\Theta)$ and $p(X|Z,\Theta)$ obtained above.

We want to maximize the log-likelihood function obtained in Question (5) above. However, since this log-likelihood function contains a sum operation Σ in the logarithmic function, it is difficult to find Θ that makes the derivative zero. Therefore, the following "Algorithm A" is used to iteratively maximize (note that here we only focus on local maximization and do not seek global maximization). In Algorithm A, for the objective function $D(\Theta)$ to be maximized, using the auxiliary function $G(\Theta, \theta)$ and the auxiliary variable θ where $D(\Theta) = \max_\theta G(\Theta, \theta)$, the operations [step 1] $\theta \leftarrow \mathrm{argmax}_{\theta'} G(\Theta, \theta')$ and [step 2] $\Theta \leftarrow \mathrm{argmax}_{\Theta'} G(\Theta', \theta)$ are iteratively repeated to calculate the local optimum of Θ. Answer the following questions.

(6) Prove that, when using above Algorithm A, Θ that is updated by repeating step 1 and step 2 always raises or stops the original objective function $D(\Theta)$.

(7) We use Jensen's inequality to obtain the auxiliary function of the log-likelihood function obtained in Question (5) above. Jensen's inequality for logarithmic functions can be written as $\log(\Sigma_i \lambda_i y_i) \geqq \Sigma_i \lambda_i \log(y_i)$ (where y_i is a positive variable, λ_i is any weight that satisfies $\Sigma_i \lambda_i = 1$ and $\lambda_i \geqq 0$). Prove this inequality.

(8) Using Jensen's inequality, design an auxiliary function for the log-likelihood function obtained in Question (5). Let λ_i be the auxiliary variable here. You can derive the auxiliary function by first dividing the individual normal distributions in the sum operation Σ by λ_i and multiplying them by λ_i in front of that.

(9) Execute Algorithm A for the auxiliary function obtained in Question (8) above. First, find λ_i that maximizes the auxiliary function as [step 1]. Next, find the parameter Θ that maximizes the auxiliary function (where the auxiliary variables are fixed) as [step 2].

Problem 2

Let us consider a dataset that consists of N data where each datum is represented in the form $x = (x_1, x_2, \dots, x_b)$ $(x_i \in \{0,1\}, 1 \leq i \leq b)$ which is a bit string of length b $(b \geq 1)$. Each datum is assigned a unique data ID (identifier) which is a distinct integer. Let's build a system that searches for data close in distance to an arbitrary input datum (query datum). During a search, the system needs to enumerate the data IDs of all data that satisfy the condition. The distance between two data is defined by the Hamming distance between bit strings. The Hamming distance between two bit strings $x = (x_1, x_2, \dots, x_b)$ and $y = (y_1, y_2, \dots, y_b)$ is defined as follows.

$$d(x,y) = \sum_{i=1}^{b} |x_i - y_i|$$

Answer the following questions.

(1) The table below shows an example of the dataset in the case of $b = 4$ and $N = 3$.

Data ID	x_1	x_2	x_3	x_4
1	0	1	1	1
2	1	0	0	1
3	0	0	1	0

Assume that a query datum $(1, 1, 0, 1)$ is given. Find the Hamming distance between the query datum and each datum.

Next, we consider a search algorithm using a lookup table. Assume that b is an even number and that the bit strings of data are uniformly distributed. In the following questions, it is not necessary to consider the time complexity of building a lookup table.

(2) We want to search for data whose bit strings are identical to a given query datum. Here, let us consider a lookup table that takes a bit string as an input and outputs a list containing the data IDs of all data that have the bit string. Answer the average time complexity and the space complexity of a search using this lookup table.

(3) We consider an algorithm as follows. We divide a bit string into two bit strings of length $b/2$. We search for candidates by using the lookup table in the same manner as Question

(2) for each divided bit string and then return the data IDs of all data matching the query datum. Answer the average time complexity and the space complexity of a search by this algorithm.

(4) By using the same data structure as Question (3), we consider an algorithm to search for data with a Hamming distance of 1 or less from a given query datum. Answer the average time complexity of a search by this algorithm.

Next, we consider the case where $2^b \gg N$. In this case, it is sometimes effective to perform a linear search by actually calculating the Hamming distance between the query datum and each datum. Therefore, let us consider designing a specialized digital circuit to compute the Hamming distance.

(5) Let us consider a 2-input 1-output digital circuit H_1 whose inputs are two 1-bit bit strings $\boldsymbol{x} = (x_1)$, $\boldsymbol{y} = (y_1)$ ($x_1, y_1 \in \{0,1\}$) and output is $z \in \{0,1\}$ which is the Hamming distance between $\boldsymbol{x}$ and $\boldsymbol{y}$. Draw a table representing the relation of $\boldsymbol{x}, \boldsymbol{y}$, and z. Also, draw H_1 by using necessary components among AND, OR, and NOT gates.

(6) Draw a 4-input 2-output digital circuit H_2 that outputs a binary representation of the Hamming distance between two 2-bit bit strings by using necessary components among AND, OR, NOT gates, and H_1.

(7) Draw an 8-input 3-output digital circuit H_4 that outputs a binary representation of the Hamming distance between two 4-bit bit strings by using necessary components among H_2, half adder *HA*, and OR gate.

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. If necessary, use examples, figures or equations.

(1) Dynamic programming
(2) Zero Moment Point (ZMP)
(3) BNF (Backus-Naur Form or Backus Normal Form)
(4) Transparent cache in wide area networks
(5) Dynamic map in self-driving car system
(6) Thread-level parallel speculative execution
(7) Procedural modeling
(8) k-nearest neighbor algorithm

このページは空白.

This page is blank.

このページは空白.

This page is blank.

このページは空白.

This page is blank.