

システム創成学専攻

2021年度 入学試験問題 論理的思考能力を見るための問題 分野1: 知能・情報科学や数理モデル

問題1-1

あなたのメールボックスには N 件のメールがある。各々のメールにはそれぞれ返信の優先度があり、この優先度は $x \in (0, 1)$ で表される。単位時間ステップ毎に、あなたは抱えているメールの中から、確率 p で「優先度が最高のもの」を、確率 $1 - p$ で「優先度を無視してランダムに選んだもの」を1件選び、そのメールに返信する。返信を済ませたメールはメールボックスから消去される。またこの1件の返信の直後に、 $(0, 1)$ 区間の一様分布に従う優先度 x を持った新たなメールが1件届く。

- (1) $N = 2$ かつ $p = 1$ の場合について考える。あなたのメールボックスにある2件のメールを、優先度の低い順に A, B と呼ぶことにしよう。つまり、これらのメールの優先度 (x_A と x_B) は、 $x_A < x_B$ である。このとき、優先度の低い方のメール A にあなたが返信するまでの時間ステップ数が t である確率を求めよ。
- (2) 次に、 $N \gg 1$ かつ $p = 1/2$ の場合について考える。上記の返信と受信の過程を繰り返した時、あなたのメールボックスの中の「メールの優先度 x の分布関数」を $f(x)$ とし、「個々のメールが届いてから返信されるまでの時間ステップ数 t の分布関数」を $g(t)$ とする。この2つの分布関数の特徴を述べよ。さらにそれらの概形も図示せよ。

問題1-2

$m + n$ 個の d 次元 ($d \geq 3$) の実数値データ $\{\mathbf{x}_1, \dots, \mathbf{x}_{m+n}\}$ が存在する。 $\{\mathbf{x}_1, \dots, \mathbf{x}_m\}$ はクラス C_1 に所属し、 $\{\mathbf{x}_{m+1}, \dots, \mathbf{x}_{m+n}\}$ はクラス C_2 に所属する。また、 m と n は $m \gg 1, n \gg 1$ とする。ここで、ある未分類のデータ \mathbf{x}_0 がどちらのクラスに所属するか、分類することを考える。以下の問に答えよ。各問では図を用いてもよい。

- (1) 分類方法として考えられるものを2つ以上挙げ、それぞれの特徴を論ぜよ。
- (2) (1)で挙げた分類方法のうちの1つについて、そのアルゴリズムを数式、疑似コード*やフローチャートなどを用いて記述せよ。
- (3) 上記(2)で説明した分類方法の精度を評価するにはどうすればよいか、その方法を述べよ。

*疑似コードとは、アルゴリズムを疑似的なプログラミング言語で記述したもの。以下は「眠くなかったら勉強する」という指示を示した例

```
if isSleepy == true:
    Go to bed
else:
    Study
```