平成 30 (2018) 年度夏入試

東京大学大学院情報理工学系研究科創造情報学専攻

創造情報学

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解答すること.
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
   その用紙で解答する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.

受験番号 ______________________________

このページは空白.

このページは空白.

第 1 問

8 つのデータを含むデータ集合 S が図 1 のように与えられている．各データは $(x_1, x_2, x_3, x_4, y) \in \{0,1\}^5$ の形式で与えられている．以下，$\mathbf{x} = (x_1, x_2, x_3, x_4)$ を $y=1$ か $y=0$ に分類するための規則を S から構成することを考える．以下の問いに答えよ．

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

図 1: データ集合

(1) 長さ n の 2 元系列 $\mathbf{z} = z_1, \ldots, z_n$ ($z_i \in \{0,1\}$, $i = 1, \ldots, n$) を符号化する際の符号長について考える．一般に，有限集合 G の 1 つの元 g を符号化するのに必要な符号長は $\log |G|$(bit) で与えられる．ただし，対数の底は 2 とする．$\mathbf{z}$ に含まれる 1 の数が k であるような長さ n の 2 元系列 $\mathbf{z}$ を，k の値も含めて符号化するための符号長は，高々

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

であることを示せ．ただし，符号化に先立って n の値は与えられているとし，符号長は非整数値を許すとする．

以下，$\mathbf{z}$ に対して式 (1) の値を $L(\mathbf{z})$ と記す．

(2) 図 1 において，$x_1 = 1$ なるデータが $y=1$ となる確率値を θ とするとき，θ の最小二乗推定値を求めよ．最小二乗推定値とは，$\sum_{t:x_1=1}(y(t)-\theta)^2$ を最小にする θ のことである．ここに，$y(t)$ は t 番目のデータの y の値を示し，$\sum_{t:x_1=1}$ は，$x_1 = 1$ なるデータ全体についての和とする．

(3) 図 1 の S のデータに対して，すべての y の値を並べて得られる 2 元系列を $\mathbf{y}$ とし，$x_i = 1$ である S の元の y の値を並べて得られる 2 元系列を $\mathbf{y}_1^{(i)}$ とし，$x_i = 0$ である S の元の y の値を並べて得られる 2 元系列を $\mathbf{y}_0^{(i)}$ とする ($i = 1, 2, 3, 4$)．例えば，$x_1 = 1$ である S の元は 1, 2, 3, 8 番目のデータなので，その対応する y を並べた 2 元系列は $\mathbf{y}_1^{(1)} = 1110$ である．このとき，$x_i = 1$ か $x_i = 0$ で S を分割することによりデータを分類することの精度指標を以下のように定める．

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

式 (2) の値が小さい程 x_i の値が y の分類に寄与していると見なす．$\Delta(i|\mathbf{y})$ を最小にする i を求めよ．ただし，以下では，最小にする i が複数存在するとき，その中からランダムに 1 つ選ぶとする．

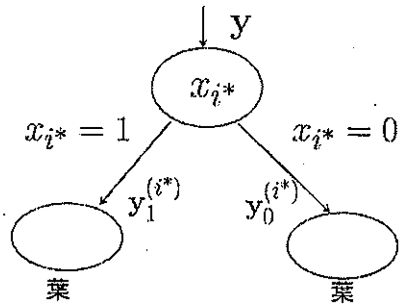

図 2: 分割木

(4) 問 (3) で求めた i を i^* とする．S は $x_{i^*}=1$ または $x_{i^*}=0$ によって分割され，$\mathbf{y}$ は $\mathbf{y}_1^{(i^*)}$ と $\mathbf{y}_0^{(i^*)}$ に分割される．これは図 2 のように木構造により表現できる．これを分割木と呼び，$\mathbf{y}_1^{(i^*)}$ と $\mathbf{y}_0^{(i^*)}$ を分割系列と呼ぶ．さらに、$\mathbf{y}_1^{(i^*)}$ と $\mathbf{y}_0^{(i^*)}$ に対し，$\Delta(i|\mathbf{y}_1^{(i^*)})$ を最小にする $i(\neq i^*)$，及び $\Delta(i|\mathbf{y}_0^{(i^*)})$ を最小化する $i(\neq i^*)$ をそれぞれ求めて同様な分割を行うとする．葉（末端）の深さ（根からその葉に辿り着くまでの分割の回数）が 2 になるか、葉に辿り着く分割系列が全て $y=1$ か全て $y=0$ からなることを停止条件として，これが満たされるまで葉の分割を繰り返すとき，最終的に得られる分割木を求めよ．

(5) 最終的に得られた分割木に対して，各葉に辿り着く分割系列において，$y=1$ の数が多ければ $y=1$ を，$y=0$ の数が多ければ $y=0$ を，同数の場合は $y=0$ か $y=1$ をランダムに木の末端（葉）に与えるとする．この分割木は新しいデータの分類予測に用いることが出来る．つまり，新たなデータの (x_1,x_2,x_3,x_4) を与えて，木を辿って到達した末端の y の値が分類予測結果となる．その際，問 (4) の停止条件を変えて，最初に与えた訓練データ S から，各葉に辿り着くデータが全て $y=1$ か全て $y=0$ になるような，より大きな分割木を作っても，必ずしも新しいデータの分類予測率を高めるとは限らない．この理由を説明せよ．

(6) 一般に、$d\geq 2$ を正整数として，$(x_1,\ldots,x_d,y)\in\{0,1\}^{d+1}$ の形式の多次元データの集合 S と，分割木 $\mathcal{T}$ が与えられたとせよ．このとき，$\mathcal{T}$ の根を共有して，$\mathcal{T}$ を葉から刈り込んで得られる部分木の全体の集合を $\mathcal{M}$ とする．$M\in\mathcal{M}$ の S に対する良さを評価するのに，木構造 M のペナルティを考慮した以下の規準を考える．

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

ここに，$N_L(M)$ は M の葉（末端）の数，$N_I(M)$ は M の内部ノードの数，C_L と C_I は与えられた正の定数とする．式 (3) の第三項の和は木の葉 $\{u\}$ 全体に渡って取られるものとし，$\mathbf{y}_u$ は葉 u に辿り着いたデータの y を並べて得られる 2 元系列であるとする．式 (3) の値が小さければ小さいほど M は良いとする．式 (3) の規準を最小化する M を $\mathcal{M}$ と S から求め，かつできるだけ時間計算量の少ないアルゴリズムを与えよ．

第2問

図1のような角度θの斜面に置かれた質量Mの台車の位置制御を考える。台車には、斜面に沿ったx軸方向へ力fを加え、移動させることができる。力fは台車を引き上げるために十分な大きさを与えることができ、台車と床の摩擦および空気抵抗は無視できるものとする。時刻tにおける力fおよび台車の位置と速度をそれぞれ$f(t)$、$x(t)$、$v(t)$と表記する。また、重力加速度の大きさをgとする。

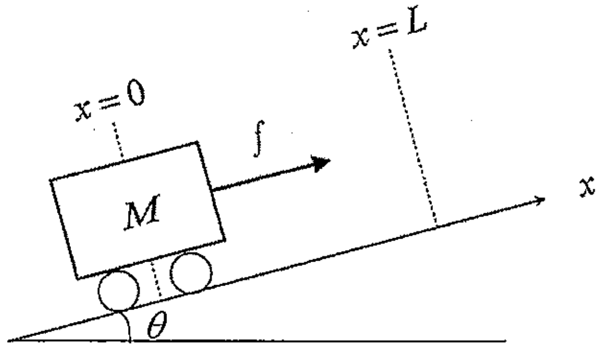

図1

時刻$t=0$において$x(0)=0$、$v(0)=0$とする。この台車を$x=L$の位置へ移動させる方法を考える。
以下の問いに答えよ。

(1) 台車を一定の力$f(t)=F\ (F>0)$で時刻t_1まで加速した時の台車の位置と速度を求めよ。
(2) (1)における時刻t_1から時刻$t_2\ (t_2 \geq t_1)$まで台車を一定の力$f(t)=-F\ (F>0)$で減速し、$x(t_2)=L$、$v(t_2)=0$としたい。この動作を実現するt_1とt_2を求めよ。

次に、目標位置である$x=L$との差に比例する力を台車に加えることを考える。すなわち、$f(t)=k_1\{L-x(t)\}$を与える。なお、k_1は正の定数とする。
(3) この場合の台車の運動方程式を表せ。
(4) $x(t)$をグラフに表せ。

次に、台車の速度に比例した力を更に加えることを考える。すなわち、
$f(t)=k_1\{L-x(t)\}-k_2 v(t)$を与える。なお、$k_1, k_2$は正の定数とする。
(5) $-k_2 v(t)$を加えることによって生じる効果と、その効果が現れる理由を説明せよ。
(6) $x(t)$が振動しないためのk_1, k_2に関する条件を求めよ。必要であれば以下の事実を用いてよい。

微分方程式 $\dfrac{d^2x}{dt^2}+a\dfrac{dx}{dt}+bx=0$ （a,b は実定数）…(A)
の一般解は、２次方程式 $r^2+ar+b=0$ …(B)
の解によって表すことができ、

1．式(B)が異なる二つの実数解 p,q を持つとき
$x=C_1e^{pt}+C_2e^{qt}$

2．式(B)が異なる二つの虚数解 $h\pm ki$ を持つとき
$x=e^{ht}(C_1\cos kt+C_2\sin kt)$

3．式(B)が重解 p を持つとき
$x=e^{pt}(C_1+C_2t)$

となる。ただし、C_1,C_2 は積分定数である。

(7) (6)で求めた条件において、$x(t)$ のグラフを表せ。

次に、目標位置との差の積分値に比例する力をさらに加えることを考える。すなわち、
$f(t)=k_1\{L-x(t)\}-k_2v(t)+k_3\int_0^t\{L-x(\tau)\}d\tau$ を与える。なお、k_1,k_2,k_3 は正の定数とする。

(8) $k_3\int_0^t\{L-x(\tau)\}d\tau$ を加えることによって生じる効果と、その効果が現れる理由を説明せよ。

第3問

以下に示す情報システムに関する8項目から<u>4項目</u>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を用いてよい．

(1) パイプライン ハザード

(2) レジスタ リネーミング

(3) カルマン フィルタ

(4) 正規文法と正規言語

(5) 公開鍵暗号方式の動作原理と認証局

(6) 巡回セールスマン問題

(7) 分割統治法

(8) ベクトル量子化

このページは空白.

このページは空白.

このページは空白.