

2027年度 大学入学共通テスト 予想問題η

試験開始の指示があるまで、この問題冊子の中を見てはいけません。

情 報 〔『情報Ⅰ』〕 (100 点 60 分)

I 注 意 事 項

- 1 この問題冊子は、35 ページあります。
- 2 解答用紙に、正しく記入・マークされていない場合は、採点できないことがあります。
- 3 試験中に問題冊子の印刷不鮮明、ページの落丁・乱丁及び解答用紙の汚れ等に気付いた場合は、手を高く挙げて監督者に知らせなさい。
- 4 問題冊子の余白等は適宜利用してよいが、どのページも切り離してはいけません。
- 5 試験終了後、問題冊子は持ち帰りなさい。

II 解 答 上 の 注 意

解答上の注意は、裏表紙に記載してあります。問題冊子を裏返して必ず読みなさい。

(下 書 き 用 紙)

情報 I の試験問題は次に続く。

(下 書 き 用 紙)

情報Ⅰの試験問題は次に続く。

情 報 I

(全 問 必 答)

第 1 問 次の問い(問 1 ～ 4)に答えよ。(配点 20)

問 1 次の問い(a ～ c)に答えよ。

- a 次の文章を読み、空欄

ア

イ

 ,

ウ

エ

 に当てはまる数字をマークせよ。

8 ビットの 2 の補数表現で -44 を表すと **11010100** となる。ビット列全体を右にずらす**シフト演算**には、空いた最上位ビットに元の最上位ビット(符号ビット)と同じ値を入れる**算術シフト**と、常に 0 を入れる**論理シフト**がある。いずれの場合も、右端からはみ出したビットは捨てる。

11010100 を 2 ビット算術右シフトした結果を 2 の補数表現として読むと

ア

イ

 となり、2 ビット論理右シフトした結果を符号なしの 2 進数として読むと

ウ

エ

 となる。

b 論理式 $\text{NOT}(A \text{ AND } B)$ と常に同じ値になる論理式として最も適当なものを，次の①～④のうちから一つ選べ。 オ

① $(\text{NOT } A) \text{ AND } (\text{NOT } B)$

② $(\text{NOT } A) \text{ OR } (\text{NOT } B)$

③ $\text{NOT}(A \text{ OR } B)$

④ $(\text{NOT } A) \text{ OR } B$

c 次の文章中の空欄 カ， キ ク ケ に当てはまる数字をマークせよ。

2進法の小数 101.011 を 10進法で表すと， カ . キ ク ケ である。

問 2 次の文章を読み，空欄 コサ，シス に当てはまる数字をマークせよ。

CPU が主記憶装置のデータを読み出すとき，主記憶装置より高速で容量の小さいキャッシュメモリに目的のデータがあればキャッシュメモリから読み出し，なければ主記憶装置から読み出す。目的のデータがキャッシュメモリにある確率をヒット率という。キャッシュメモリのアクセス時間が 10 ナノ秒，主記憶装置のアクセス時間が 70 ナノ秒のとき，平均のアクセス時間(実効アクセス時間)は次の式で求められる。

$$\text{実効アクセス時間} = h \times 10 + (1 - h) \times 70 \quad (h \text{ はヒット率})$$

ヒット率が 90% のとき，実効アクセス時間は コサ ナノ秒である。また，実効アクセス時間を 25 ナノ秒以下にするためには，ヒット率を シス % 以上にする必要がある。

問 3 次の文章を読み，空欄 セ，ソタ，チツ，テ に当てはまる数字をマークせよ。また，空欄 ト に入れるのに最も適当なものを，後の解答群のうちから一つ選べ。

仮想記憶方式のコンピュータでは，プログラムをページと呼ぶ単位に分けて管理し，必要なページだけを主記憶装置のページ枠に読み込む。参照したいページがページ枠にない場合をページフォールトと呼び，そのページを補助記憶装置から読み込む。ページ枠に空きがない場合は，ページ枠のいずれかのページを追い出す。追い出すページの選び方として，次の二つの方式を考える。

FIFO 最も早く読み込まれたページを追い出す。

LRU 最も長い間参照されていないページを追い出す。

最初はすべてのページ枠が空であるとし、ページを次の順序で参照する。

1, 2, 3, 4, 1, 2, 5, 1, 2, 3, 4, 5

FIFO では、ページ枠が 3 個の場合のページフォールトの回数は **セ** 回であり、ページ枠を 4 個に増やすと **ソ** **タ** 回となる。一方、LRU では、ページ枠が 3 個の場合は **チ** **ツ** 回、4 個の場合は **テ** 回となる。このことから、**ト** ことがわかる。

ト の解答群

- ① どちらの方式でも、ページ枠を増やせば必ずページフォールトの回数は減る
- ② LRU では、ページ枠を増やすとページフォールトの回数が必ず増える
- ③ ページフォールトの回数は、ページ枠の数によらず常に同じである
- ④ FIFO では、ページ枠を増やしてもページフォールトの回数が増える場合がある

問 4 次の文章を読み、空欄 **ナ**，**ニ****ヌ** に当てはまる数字をマークせよ。また、空欄 **ネ** に入れるのに最も適当なものを、後の解答群のうちから一つ選べ。

ある窓口では、客が平均すると 1 時間あたり 12 人、でたらめな間隔で到着し、1 人の対応には平均 4 分かかる。窓口は一つであり、客は到着順に対応を受ける。このような窓口では、窓口が対応中である時間の割合(利用率)を ρ とすると、客が列に並んで待つ時間の平均は、次の式で近似できることが知られている。

$$\text{平均待ち時間} = \rho \div (1 - \rho) \times (1 \text{ 人あたりの平均対応時間})$$

この窓口の利用率は 0. **ナ** であり、平均待ち時間は **ニ****ヌ** 分となる。また、到着する客が 1 時間あたり 13.5 人に増えると、**ネ**。

ネ の解答群

- ① 客は 1.125 倍に増えるだけなので、平均待ち時間もほぼ 1.125 倍の約 18 分になる
- ② 利用率が 1 に近づくため、平均待ち時間は 2 倍以上の 36 分になる
- ③ 利用率は変わらないので、平均待ち時間も変わらない
- ④ 利用率が 1 を超えるので、列は限りなく長くなる

第2問 次の問い(A・B)に答えよ。(配点 30)

A 次の文章を読み、後の問い(問1～6)に答えよ。

X社は、自社のWebサイトとメールサーバをインターネットに公開している。図1は、X社のネットワーク構成である。公開するサーバは、社内LANとは別の区画(DMZ)に置かれ、インターネット、DMZ、社内LANの間の通信は、すべてファイアウォール(FW)を通過する。

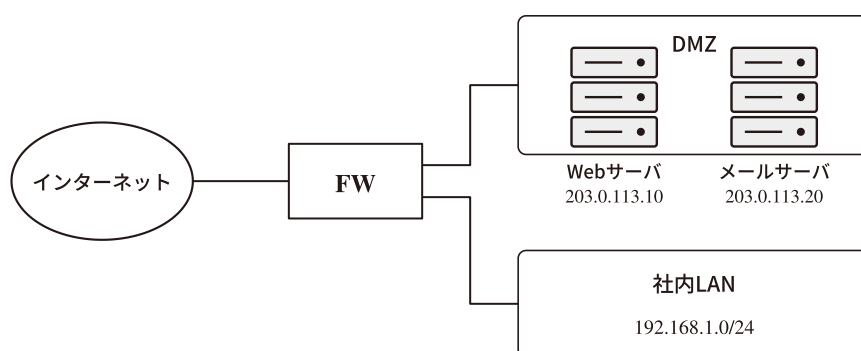


図1 X社のネットワーク構成

FWは、通過しようとするパケットを表1のルールに上から順に照らし合わせ、最初に条件がすべて一致したルールの動作(許可または拒否)を適用する。表1の「任意」はどの値とも一致することを表す。

表1 FW のルール

番号	送信元	宛先	プロトコル	宛先ポート	動作
1	任意	203.0.113.10	TCP	443	許可
2	任意	203.0.113.10	TCP	80	許可
3	192.168.1.0/24	任意	UDP	53	許可
4	任意	203.0.113.20	TCP	25	許可
5	192.168.1.0/24	任意	TCP	443	許可
6	任意	任意	TCP	22	拒否
7	任意	任意	任意	任意	拒否

表2 FW を通過しようとしたパケット

パケット	送信元	宛先	プロトコル	宛先ポート
①	198.51.100.5	203.0.113.10	TCP	443
②	198.51.100.5	203.0.113.10	TCP	22
③	192.168.1.15	198.51.100.80	TCP	443
④	192.168.1.15	198.51.100.80	TCP	80
⑤	198.51.100.7	203.0.113.20	TCP	25
⑥	192.168.2.3	8.8.8.8	UDP	53
⑦	192.168.1.20	8.8.8.8	UDP	53

問1 次の文章中の空欄 **ア**， **イ**， **ウ** に当てはまる数字をマークせよ。

表2のパケット②に適用されるのは表1の番号 **ア** のルールであり、パケット④に適用されるのは番号 **イ** のルールである。また、表2の7個のパケットのうち、FWを通過できるのは **ウ** 個である。

問 2 X 社の管理者は，社内 LAN の PC から Web サーバ(203.0.113.10)に SSH (TCP の 22 番ポート)で接続して保守作業を行いたい。他の通信の扱いを変えずにこれを実現するための表 1 の変更として最も適当なものを，次の①～③のうちから一つ選べ。

エ

- ① 同じ内容のルールを，番号 7 のルールの直後に追加する。
- ② 番号 6 のルールを削除する。
- ③ 番号 7 のルールの動作を「許可」に変更する。
- ④ 「送信元：192.168.1.0/24，宛先：203.0.113.10，プロトコル：TCP，宛先ポート：22，動作：許可」というルールを，番号 6 のルールの直前に追加する。

問 3 Web サーバのログイン機能では、入力された利用者 ID とパスワードを使って、次のような SQL 文を組み立ててデータベースに問い合わせている。ここで、「--」以降はコメントとして無視される。

```
SELECT * FROM users
WHERE id = '（入力した ID）'
AND pw = '（入力したパスワード）'
```

利用者 ID として「admin'--」と入力し、パスワードには何も入力しなかった場合に起こりうることで最も適当なものを **オ**、この攻撃への対策として最も適当なものを **カ** から一つずつ選べ。

オ の解答群

- ④ 入力の文字数が多すぎるため、サーバが停止する
- ① 利用者 ID に記号が含まれるので、必ずエラーになり誰もログインできなくなる
- ② パスワードの条件がコメントとして無視され、admin のパスワードを知らなくても admin としてログインできてしまう
- ③ パスワードが空なので、通信が暗号化されずに送信される

カ の解答群

- ④ Web サーバとの通信を TLS で暗号化する
- ① SQL 文を組み立てるときにプレースホルダを用い、入力値を SQL 文の一部として解釈させないようにする
- ② パスワードの最低文字数を長くする
- ③ FW で宛先ポート 443 以外を拒否する

問 4 次の文章を読み，空欄 **キ** ・ **ク** に入れるのに最も適当なものを，後の解答群のうちから一つずつ選べ。

X 社のサーバでは，パスワードをそのまま保存せず，利用者ごとに異なるランダムな文字列(ソルト)をパスワードに付け加えてからハッシュ値を求めて保存している。ソルトを用いる主な目的は， **キ** ことである。

さらに，ハッシュ値を求める計算を 10000 回繰り返す **ストレッチング** も行っている。ハッシュ値の 1 回の計算に 1 マイクロ秒(100 万分の 1 秒)かかるとすると，攻撃者がパスワードの候補 1 億個を一つずつ試すのにかかる時間は，ストレッチングを行わない場合の 100 秒に対し，およそ **ク** となる。

キ の解答群

- ④ ハッシュ値から元のパスワードを計算で求められるようにする
- ① パスワードの文字数を自動的に増やして，利用者が覚えやすくする
- ② ログインの通信を暗号化する
- ③ 同じパスワードを使っている利用者がいても保存されるハッシュ値が異なるようにし，あらかじめ計算したハッシュ値の一覧による解読を困難にする

ク の解答群

- ④ 約 1.2 日 ① 約 12 日 ② 約 120 日 ③ 約 1200 日

問 5 Web サイトの問合せフォームでは，入力された文字列をそのまま確認画面の HTML に埋め込んで表示している。このとき，攻撃者がスクリプトを含む文字列を入力させるように仕向けたリンクを利用者に開かせると，利用者のブラウザでそのスクリプトが実行されるおそれがある。この攻撃と対策の組合せとして最も適当なものを，次の①～③のうちから一つ選べ。

ケ

- ① クロスサイトスクリプティング：Web サーバの前に FW を設置する。
- ② DoS 攻撃：入力された文字列をハッシュ値に変換して保存する。
- ③ クロスサイトスクリプティング：入力された文字列のうち，「<」や「>」など HTML で特別な意味をもつ文字を，出力するときに別の表現に置き換える(エスケープする)。
- ④ SQL インジェクション：パスワードにソルトを付け加える。

問 6 X 社の Web サイトでは，HTTPS で通信するためにサーバ証明書を用いている。利用者のブラウザが Web サーバから受け取ったサーバ証明書を検証することで確認できることとして最も適切なものを，次の①～③のうちから一つ選べ。

- ① Web サーバのプログラムに脆弱性がないこと
- ② Web サイトに掲載されている内容が正確であること
- ③ 通信相手の Web サーバが，アクセスしようとしたドメイン名の正当なサーバであること
- ④ Web サーバに保存されている利用者のパスワードが暗号化されていること

B 次の文章を読み、後の問い(問 1 ～ 5)に答えよ。

X 社は、社内 LAN を部署ごとのネットワーク(サブネット)に分けることにした。社内 LAN に割り当てられたアドレスは 192.168.10.0/24 である。各部署で必要な機器の数(IP アドレスの数)は、営業部 100、開発部 50、総務部 20、会議室 10 である。各サブネットでは、ホスト部がすべて 0 のアドレスとすべて 1 のアドレス(ブロードキャストアドレス)は機器に割り当てられない。

必要な機器の数を収容できる最小の大きさのサブネットを、必要な数が多い部署から順に、192.168.10.0 から隙間なく割り当てていく。例えば、営業部には 192.168.10.0/25 が割り当てられ、機器に割り当てられるのは 192.168.10.1 ～ 192.168.10.126 の 126 個である。

問 1 次の文章中の空欄

サ	シ	ス
---	---	---

、

セ	ソ
---	---

、

タ	チ	ツ
---	---	---

 に当てはまる数字をマークせよ。

総務部に割り当てられるサブネットは次のとおりである。

192.168.10.

サ	シ	ス
---	---	---

 /

セ	ソ
---	---

また、会議室のサブネットのブロードキャストアドレスは次のとおりである。

192.168.10.

タ	チ	ツ
---	---	---

問 2 次の文章を読み、空欄 **テ** ～ **ナ** に入れるのに最も適当なものを、後の解答群のうちから一つずつ選べ。ただし、同じものを繰り返し選んでもよい。

ルータは、宛先 IP アドレスと表 3 の経路表を照らし合わせ、宛先が含まれるネットワークのうち、ネットワーク部が最も長い(プレフィックス長が最も大きい)行の転送先にパケットを転送する。0.0.0.0/0 はすべてのアドレスを含む。

表 3 ルータの経路表

宛先ネットワーク	転送先
10.0.0.0/8	R1
10.1.0.0/16	R2
10.1.2.0/24	R3
10.1.2.128/25	R4
0.0.0.0/0	R5

このとき、宛先が次の(a)～(c)のパケットは、それぞれ **テ** ～ **ナ** に転送される。

(a) 宛先 10.1.2.200 → **テ**

(b) 宛先 10.1.3.5 → **ト**

(c) 宛先 172.16.0.1 → **ナ**

— **テ** ～ **ナ** の解答群 —

① R1	② R2	③ R3	④ R4	⑤ R5
------	------	------	------	------

問 3 次の文章中の空欄 に当てはまる数字をマークせよ。

X 社の他の拠点では，192.168.8.0/24，192.168.9.0/24，192.168.10.0/24，192.168.11.0/24 の四つのネットワークを使っている。これら四つをまとめて一つの経路として経路表に登録すると，192.168.8.0/ と表せる。

問 4 次の文章中の空欄 に当てはまる数字をマークせよ。

拠点間は通信速度 100Mbps の回線で結ばれており，実際に利用できるのは通信速度の 80% である。このとき，600MB のファイルを転送するのにかかる時間は 秒である。ただし，1MB = 1000000 バイトとし，転送以外の時間は考えないものとする。

問 5 次の文章を読み、空欄 **ハ** ・ **ヒ** に入れるのに最も適当なものを、後の解答群のうちから一つずつ選べ。

X 社の FW は、社内 LAN の PC がインターネット上のサーバと通信するとき、送信元の IP アドレスとポート番号を、FW のインターネット側の IP アドレス 203.0.113.1 と、FW が選んだポート番号に変換して送信する (NAPT)。FW は変換の対応を表 4 のように記録しておき、サーバから戻ってきたパケットの宛先を、この表に従って元に戻す。

表 4 FW が記録している変換の対応

社内側	インターネット側
192.168.1.15 : 50001	203.0.113.1 : 60001
192.168.1.20 : 50001	203.0.113.1 : 60002
192.168.1.15 : 50002	203.0.113.1 : 60003

(注)「IP アドレス：ポート番号」の形式で表している。

インターネット上のサーバから、宛先が 203.0.113.1 : 60002 のパケットが届いた場合、FW はこのパケットを **ハ** に転送する。また、NAPT を用いる利点として最も適当なものは **ヒ** である。

ハ の解答群	
① 192.168.1.15 : 50001	① 192.168.1.20 : 50001
② 192.168.1.15 : 50002	③ 203.0.113.1 : 60002

ヒ の解答群

- ④ 通信の内容が自動的に暗号化される
- ① 社内の PC どうしの通信速度が速くなる
- ② インターネット上のサーバの IP アドレスを変更できる
- ③ 一つのグローバル IP アドレスを，社内の複数の PC で同時に共有してインターネットと通信できる

第3問 次の文章を読み、後の問い(問1～7)に答えよ。(配点 25)

データの探索を効率よく行うための構造として、**二分探索木**がある。二分探索木では、各データ(以下、**節**と呼ぶ。)は最大二つの子をもつ。どの節についても、左側の子とその下に連なるすべての節の値はその節の値より小さく、右側の子とその下に連なるすべての節の値はその節の値より大きい。

Kさんは、二分探索木を三つの配列で表すことにした。配列 **Atai** には値を追加した順に格納し、その添字を節の番号とする。配列 **Hidari** と **Migi** には、それぞれ左の子と右の子の節の番号を格納し、子がない場合は0とする。配列の添字は1から始まり、番号1の節を**根**と呼ぶ。

最初の値を根とし、2番目以降の値は、根から順に、追加する値が節の値より小さければ左の子へ、大きければ右の子へ進み、進む先がなくなった(0である)ところに新しい節として追加する。図1は、50, 30, 70, 20, 40, 60, 80の順に追加した二分探索木である。なお、同じ値が追加されることはないものとする。

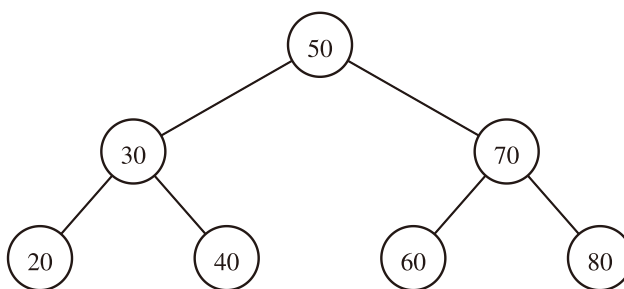


図1 7個の値を追加した二分探索木

問 1 次の文章中の空欄 、、、 に当てはまる数字をマークせよ。

図 1 の二分探索木に、続けて 35, 45, 65 の順に値を追加した。このとき、**Atai**[5] は 40 であり、**Hidari**[5] は , **Migi**[5] は となる。根から最も遠い節までの節の数(根と最も遠い節を含む)を高さと呼ぶと、この二分探索木の高さは である。また、この二分探索木で 55 を探するとき、根から順に値を比較していくと、 回比較したところで 55 がないことがわかる。

問 2 次の文章を読み、空欄 ～ に入れるのに最も適当なものを、後の解答群のうちから一つずつ選べ。

K さんは、値を順に追加して二分探索木を作るプログラムを作成した(図 2)。変数 **p** は現在調べている節の番号を表し、変数 **kettei** は追加する場所が決まったら 1 にする。

```

(01) Atai = [50, 30, 70, 20, 40, 60, 80, 35, 45, 65]
(02) Hidari = [0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
(03) Migi    = [0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
(04) i を 2 から 要素数(Atai) まで 1 ずつ増やしながら繰り返す:
(05) |   p = 1
(06) |   kettei = 0
(07) |   kettei == 0 の間繰り返す:
(08) |   |   もし キ ならば:
(09) |   |   |   もし Hidari[p] == 0 ならば:
(10) |   |   |   |   Hidari[p] = オ
(11) |   |   |   |   kettei = 1
(12) |   |   |   そうでなければ:
(13) |   |   |   |   p = カ
(14) |   |   |   そうでなければ:
(15) |   |   |   もし Migi[p] == 0 ならば:
(16) |   |   |   |   Migi[p] = オ
(17) |   |   |   |   kettei = 1
(18) |   |   |   そうでなければ:
(19) |   |   |   |   p = Migi[p]

```

図2 二分探索木を作るプログラム

オ

の解答群

- ① p ② i ③ Atai[i] ④ Atai[p]

カ

の解答群

- ① Migi[p] ② Hidari[i] ③ p + 1 ④ Hidari[p]

キ

の解答群

- ① Atai[i] > Atai[p] ② i < p
③ Atai[i] < Atai[p] ④ Atai[p] < Hidari[p]

問 3 次の文章中の空欄 に当てはまる数字をマークせよ。

問 1 の 10 個の値からなる二分探索木で、格納されている 10 個の値をそれぞれ一回ずつ探すとき、根から順に比較して目的の値が見つかるまでの比較の回数の合計は 回である。

問 4 次の文章を読み，空欄 ・ に入れるのに最も適当なものを，後の解答群のうちから一つずつ選べ。また，空欄 に当てはまる数字をマークせよ。

二分探索木の節の値を小さい順に表示するには，各節について「左の子以下の節をすべて表示し，その節を表示し，右の子以下の節をすべて表示する」という順で節をたどればよい。K さんは，この順序でたどるプログラムを，配列 **Stack** を用いて作成した(図 3)。**Stack** は，後でたどる節の番号を一時的に積み上げておくための配列であり，変数 **sp** は積み上げられている番号の個数を表す。また，「**or**」は左右の式のいずれか一方でも真のときに真となる。

```
(01) Stack = [0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
(02) sp = 0
(03) p = 1
(04) p != 0 or sp > 0 の間繰り返す:
(05) | p != 0 の間繰り返す:
(06) | | sp = sp + 1
(07) | | Stack[sp] = p
(08) | | p = 
(09) | p = Stack[sp]
(10) | sp = sp - 1
(11) | 表示する (Atai[p])
(12) | p = 
```

図 3 値を小さい順に表示するプログラム

問 1 の二分探索木で図 3 のプログラムを実行すると，**sp** の値は最大で になる。

コ・サの解答群

- | | | |
|-------------|-----------|-------------|
| ① Hidari[p] | ① Migi[p] | ② Stack[sp] |
| ③ Atai[p] | ④ p + 1 | ⑤ 0 |

問 5 次の文章を読み、空欄 **ス** **セ** に当てはまる数字をマークせよ。また、空欄 **ソ** に入れるのに最も適当なものを、後の解答群のうちから一つ選べ。

問 1 と同じ 10 個の値を、小さい順(20, 30, 35, …, 80)に追加して二分探索木を作ると、高さは **ス** **セ** となる。このように、二分探索木は値を追加する順序によって形が偏り、探索の効率が悪くなることがある。これを防ぐ方法として最も適当なものは **ソ** である。

ソの解答群

- ① 値を必ず小さい順に並べてから追加する
- ① 根の値を常に最小の値にする
- ② 値を追加するたびに、左右の高さの差が大きくなりすぎないように木の形を組み替える
- ③ 配列 **Hidari** と **Migi** の要素数を 2 倍にする

問 6 次の文章を読み，空欄 **タ** に入れるのに最も適当なものを，後の解答群のうちから一つ選べ。また，空欄 **チ** に当てはまる数字をマークせよ。

二分探索木から，子を二つもつ節を削除する場合には，その節の右の子以下の節のうち最も小さい値をもつ節の値を削除する節の位置に移し，その最も小さい値をもつ節を代わりに取り除く。こうすると，二分探索木の性質を保つことができる。

問 1 の 10 個の値からなる二分探索木から 30 を削除すると，30 の位置には **タ** が移る。削除した後の二分探索木で 45 を探するとき，根から順に値を比較すると，**チ** 回目の比較で 45 が見つかる。

— タ の解答群 —				
① 20	② 35	③ 40	④ 45	⑤ 50

問 7 次の文章を読み，空欄 に当てはまる数字をマークせよ。また，空欄 に入れるのに最も適当なものを，後の解答群のうちから一つ選べ。

1023 個の値を格納した二分探索木が，根から見て左右に最も偏りなく作られている(どの節も，空の場所がない限り子を二つもつ)とき，高さは である。このとき，格納されている値を探すのに必要な比較の回数は最大でも 回である。これに対して，1023 個の値を配列の先頭から順に調べる方法では，最大で 1023 回の比較が必要になる。値の個数を n とすると，偏りのない二分探索木での最大の比較の回数は，およそ に比例して増える。

—— の解答群 ——

①	n	②	n の 2 乗	③	$\log_2 n$	④	2 の n 乗
---	-----	---	-----------	---	------------	---	-----------

第4問 次の文章を読み、後の問い(問1～6)に答えよ。(配点 25)

Y社では、受信したメールが迷惑メールかどうかを判定するプログラムを開発している。このプログラムは、機械学習によって作成したモデルを用いて、各メールについて迷惑メールらしさを表す0以上1未満の**スコア**を求め、スコアがあらかじめ決めた値(**しきい値**)以上であれば迷惑メールと判定する。

モデルの性能を評価するため、迷惑メールかどうかわかっている1000通のメール(正当なメール800通, 迷惑メール200通)のスコアを求め、図1にまとめた。図1の各階級は、例えば0.5以上0.6未満のように区切っている。

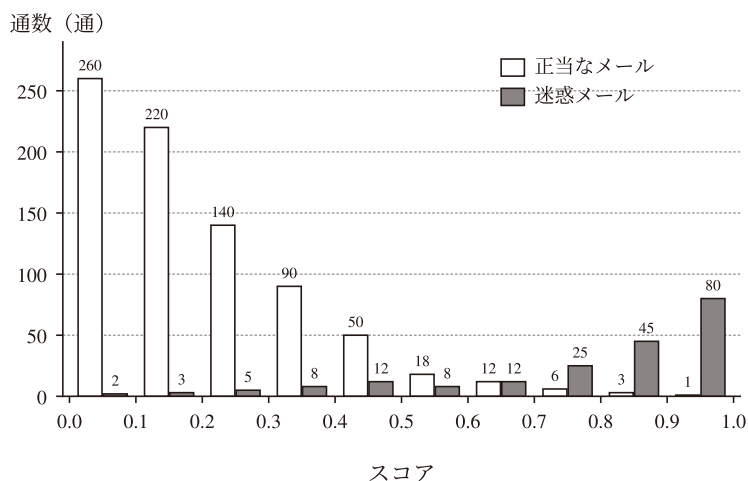


図1 スコアの度数分布(1000通)

問 1 次の文章を読み、空欄 アイウ，エオ に当てはまる数字をマークせよ。

判定の結果は、表 1 のように 4 種類に分類できる。しきい値を 0.5 としたとき、表 1 の TP は アイウ 通， FP は エオ 通である。

表 1 判定結果の分類

	迷惑メールと判定	正当なメールと判定
実際は迷惑メール	TP	FN
実際は正当なメール	FP	TN

問 2 次の文章を読み、空欄 カキ，クケ，コサ に当てはまる数字をマークせよ。

モデルの評価には、次の三つの指標がよく用いられる。

- ・ **正解率**：全体のうち判定が正しかった割合 $(TP + TN) \div \text{全体}$
- ・ **適合率**：迷惑メールと判定したもののうち、実際に迷惑メールだった割合 $TP \div (TP + FP)$
- ・ **再現率**：実際の迷惑メールのうち、迷惑メールと判定できた割合 $TP \div (TP + FN)$

しきい値を 0.5 としたとき、小数第 3 位を四捨五入すると、適合率は 0. カキ，再現率は 0. クケ，正解率は 0. コサ である。

問 3 次の文章を読み、空欄 シ に入れるのに最も適当なものを、後の解答群のうちから一つ選べ。

しきい値を 0.7 に変更すると、適合率は 0.94、再現率は 0.75 になる。Y 社では、正当なメールを迷惑メールと誤って判定すると、重要な連絡を見落とす被害が大きいと考えている。このとき、しきい値の選び方として最も適当なものは シ である。

シ の解答群

- ④ 再現率を高くするため、しきい値を 0.5 より小さくする
- ① 正当なメールを迷惑メールと判定する件数を減らすため、適合率が高くなるようにしきい値を 0.7 にする
- ② しきい値を変えても *FP* は変わらないので、どちらでもよい
- ③ 迷惑メールを一つも見逃さないように、しきい値を 0 にする

問 4 次の文章を読み、空欄 **ス** **セ** に当てはまる数字をマークせよ。また、空欄 **ソ** に入れるのに最も適当なものを、後の解答群のうちから一つ選べ。

スコアを使わずに、すべてのメールを正当なメールと判定するプログラムの正解率は 0. **ス** **セ** である。このことから、 **ソ** 。

ソ の解答群

- ④ すべてのメールを正当なメールと判定するプログラムの方が、モデルを使うよりも性能がよい
- ① 迷惑メールの割合が少ない場合、正解率だけでは判定の性能を正しく評価できないので、適合率や再現率も合わせて確認する必要がある
- ② 正解率が 0.8 以上であれば、迷惑メールの判定には十分な性能である
- ③ 再現率は、迷惑メールの割合に関係なく常に正解率と等しくなる

問 5 次の文章を読み、空欄 **タ** に当てはまる数字をマークせよ。また、空欄 **チ** ・ **ツ** に入れるのに最も適当なものを、後の解答群のうちから一つずつ選べ。

Y 社では、モデルの複雑さ(調整できるパラメータの多さ)を 10 段階で変えて、モデルを作り直した。モデルの作成に用いたメール(学習用データ)とは別に、作成に用いなかったメール(検証用データ)を用意し、それぞれでの正解率を図 2 にまとめた。

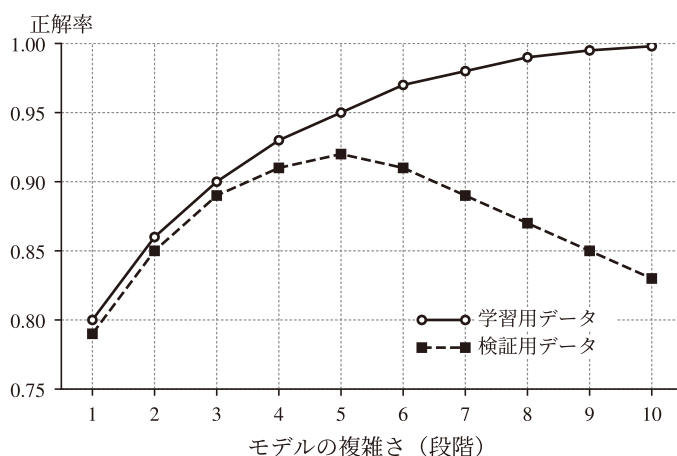


図 2 モデルの複雑さと正解率

新しいメールの判定に用いるモデルとしては、複雑さが **タ** 段階のものが最も適当であると考えられる。複雑さを 6 段階以上になると、学習用データでの正解率は上がり続けるが検証用データでの正解率は下がっている。このような状態を **チ** という。また、データを複数のグループに分け、検証に使うグループを入れ替えながら学習と検証を繰り返して性能を評価する方法を **ツ** という。

チの解答群

- ① 未学習 ② 正規化 ③ 過学習 ④ 標本化

ツの解答群

- ① 主成分分析 ② 交差検証
③ 仮説検定 ④ クラスタリング

問 6 次の文章を読み、空欄 **テ** **ト** に当てはまる数字をマークせよ。また、空欄 **ナ** に入れるのに最も適当なものを、後の解答群のうちから一つ選べ。

適合率と再現率は、一方を高くしようとするとは方が低くなりやすい。そこで、両方をまとめた指標として、次の式で求める **F 値** が用いられる。

$$F \text{ 値} = 2 \times (\text{適合率}) \times (\text{再現率}) \div \{(\text{適合率}) + (\text{再現率})\}$$

しきい値を 0.5 としたときの **F 値** は、小数第 3 位を四捨五入すると 0. **テ** **ト** である。**F 値** は適合率と再現率の **ナ** であり、どちらか一方が極端に低いと、**F 値** も低くなる。

ナの解答群

- ① 加重平均 ② 中央値 ③ 最大値 ④ 調和平均

(下 書 き 用 紙)

II 解答上の注意

- 1 解答は、解答用紙の問題番号に対応した解答欄にマークしなさい。例えば、第2問の **ア** と表示のある問いに対して③と解答する場合は、次の例のように問題番号 **2** の解答記号 **ア** の解答欄の③にマークしなさい。

例1

2	解 答 欄															
	0	1	2	3	4	5	6	7	8	9	a	b	c	d	e	f
ア	①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	Ⓕ

- 2 この試験では、選択肢から選んで答える問題と、解答欄の数字(0～9)又は文字(a～f)から選んで答える問題があります。各問題の指示に従って解答しなさい。
- 3 問題の文中の **イ**，**ウ** **エ** などの **□** に、数字(0～9)又は文字(a～f)を入れるよう指示された場合、**イ**，**ウ**，**エ**…の一つ一つは、これらのいずれか一つに対応します。それらを解答用紙の**イ**，**ウ**，**エ**…で示された解答欄にマークして答えなさい。

例2 **ウ** **エ** に 38 と答えたいとき

ウ	①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	Ⓕ
エ	①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	Ⓕ

- 4 同一の問題文中に **ア**，**ウ** **エ** などが2度以上現れる場合、原則として、2度目以降は、**ア**，**ウ** **エ** のように細字で表記します。

※この問題冊子は、大学入学共通テスト「情報Ⅰ」の過去の出題傾向をもとに作成した予想問題であり、独立行政法人大学入試センターが作成・公表したものではありません。また、問題中のデータは架空のものです。