

検 査 Ⅲ 情 報

(解答上の注意) 解答は、すべて解答用紙に記入すること。

1 W高校で、教職員向けに情報セキュリティに関する研修を行った。次の各問いに答えなさい。

- (1) 次の文章は Windows OS を利用しているパソコンについて、サポート詐欺に対する対処方法を、情報セキュリティ白書 2024 (独立行政法人情報処理推進機構 2024 年 7 月 30 日発行) の第 1 章「1.2.7 個人を狙う様々な騙しと悪用の手口 (1) 偽のセキュリティ警告 (サポート詐欺) (b) 対処」から抜粋したものである。①～⑤に入る適切な語句を答えなさい。

パソコンの警告画面については、Web ブラウザーを閉じるだけでよい。画面を閉じるには、「ESC」キーを長押しして Web ブラウザーのフルスクリーン表示を解除した上で画面右上に現れた「×」(閉じる) ボタンをクリック、または「(①)」「(②)」「(③)」キーを同時に押してからパソコンの再起動を行う。

(中略)

パソコンに遠隔操作ソフトウェアをインストールさせられた場合は、Windows の「システムの (④)」機能を使用して、当該ソフトウェアをインストールする前の状態にシステムを戻すことを推奨する。遠隔操作の及ぼす影響について判断できないため、システムの (④) ができない場合は、パソコンの (⑤) を推奨する。

- (2) 情報セキュリティ白書 2024 の第 1 章では、個人を狙う様々な手口が次の①～④のように解説されている。それぞれ何を解説しているのか、適切な語句を、下のア～エの中から選び、記号で答えなさい。

- ① 巧妙な騙しの手口を駆使した偽のメールを企業・組織に送り付け、役職員を騙して送金取引に関わる資金を詐取するなどの金銭被害をもたらすサイバー攻撃の一種である。偽のメールを送るための前段階として、企業の役職員や取引先のメールアカウント情報を狙うケースもあり、フィッシング攻撃や情報を窃取するウイルスを使用することもある。
- ② 「ransom」(身代金)と「software」(ソフトウェア)を組み合わせた造語である。パソコンやサーバーなどのシステムをロックすることや、システムに保存されているファイルを暗号化することにより、機器を使用不能にするウイルスの総称として用いられる。
- ③ Web サーバーなどの攻撃対象に対して、複数の送信元から同時に大量のパケットや問い合わせを送信することで、攻撃対象のリソースに負荷をかけ、サービス運用を妨害する攻撃である。
- ④ ある特定の企業・組織や業界などを狙って行われるサイバー攻撃の一種である。フィッシングメールやウイルスメールを不特定多数の相手に無差別に送り付ける攻撃とは異なり、標的とする特定の企業・組織や業界が持つ機密情報の窃取など明確な目的をもって行われる。

DDoS

ア ランサムウェア イ 標的型攻撃 ウ ビジネスメール詐欺 エ ~~DDos~~ 攻撃

※選択肢に DDos と誤記がありました。(2)①～④については全員を正答とします。

検査Ⅲ 情 報

- (3) W高校の情報セキュリティ対策委員会が電子メールについて調査を行い、以下のレポートをまとめた。①～③に入る適切な数値を答えなさい。ただし、③に入る数値は小数第1位を四捨五入して答えなさい。

レポート

調査対象

調査期間

調査方法

集計結果

W高校の学校代表メールアドレスで受信したすべての電子メール

2025 年X月の1ヶ月間

(i) すべての電子メールについて委員会が迷惑メールか非迷惑メールかを判定

(ii) (i) の判定結果ごとに、電子メールソフトの自動振り分け機能では迷惑メール、非迷惑メールのいずれに判定されているか調査

(i) により迷惑メールは受信したすべての電子メールの0.5%であった。

(i) の判定結果ごとの、(ii) の判定結果の割合は次の表のようになった。

		(ii)	
		迷惑メールと判定	非迷惑メールと判定
(i)	迷惑メール	97%	3%
	非迷惑メール	2%	98%

分 析

「条件付き確率」を用いて、電子メールソフトの自動振り分け機能で迷惑メールと判定されたメールが、委員会でも迷惑メールと判定されている確率を求めてみた。

あるメールが、委員会により迷惑メールと判定される事象をA、電子メールソフトの迷惑メールの自動振り分け機能により迷惑メールと判定される事象をBとする。

$P(A \cap B) = P(A) \times P_A(B) = 0.005 \times 0.97 =$

①

$P(\bar{A} \cap B) = P(\bar{A}) \times P_{\bar{A}}(B) = 0.995 \times 0.02 = 0.0199$

A ∩ B と $\bar{A} \cap B$ は互いに排反であるから

$P(B) = P(A \cap B) + P(\bar{A} \cap B) =$

①

+

0.0199 =

②

よって、求める確率は $P_B(A) = P(A \cap B) \div P(B) =$

①

÷

②

= 0.1959...

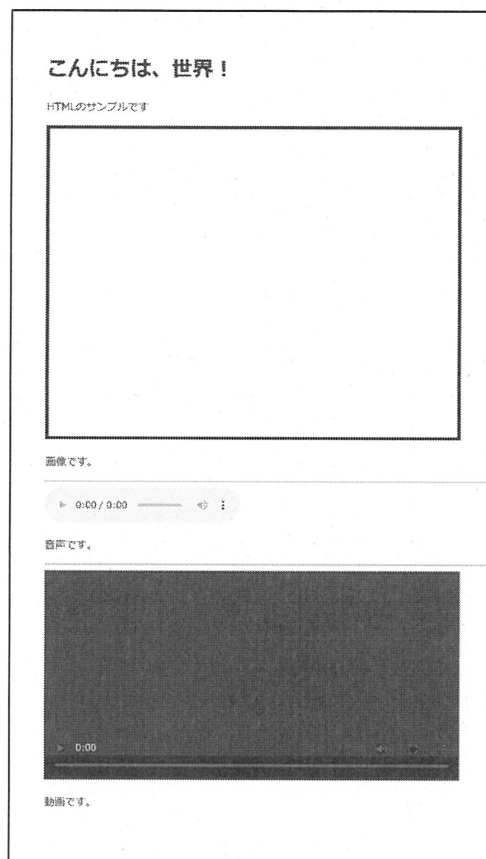
この計算結果では、迷惑メールの自動振り分け機能を持つ電子メールソフトで迷惑メールと判定されても、そのうちの約 (③) % は非迷惑メールである可能性を示している。

検査Ⅲ 情報 報

2 情報のデジタル化の教材として、下記のような HTML ファイルを作成したい。次の各問いに答えなさい。

行	ソース
1	<!DOCTYPE html>
2	<html lang="ja">
3	<head>
4	<meta charset="UTF-8">
5	<meta name="viewport" content=
6	"width=device-width, initial-scale=1.0">
7	<title>HTML のサンプルです</title>
8	</head>
9	<body>
10	こんにちは、世界！
11	HTML のサンプルです
12	
14	画像です。
15	<hr>
16	<audio controls>
17	<source src="example.mp3"
18	type="audio/mpeg">
19	</audio>
20	音声です。
21	<hr>
22	<video controls width="640">
23	<source src="example.mp4"
24	type="video/mp4">
25	</video>
26	動画です。
27	</body>
28	</html>

表示例



検査Ⅲ 情 報

- (1) 4行目は文字コードを指定している。「UTF-8」をはじめとした「Unicode」が策定された目的を答えなさい。
- (2) 11行目の example.bmp は画素数が 640×480 のビットマップ形式である。このファイルが 1画素あたり 8ビットのモノクロ 256階調であった場合と、1画素あたり 24ビットのフルカラー画像であった場合を比較するとデータサイズの差は何 KB となるか答えなさい。なお、1KB は 1,024B とする。
- (3) 15行目は音声ファイルを指定しているが、音楽 CD を素材にしている。音楽 CD の標準化周波数は 44.1kHz であるが、これは人間の可聴周波数の上限が 20kHz であることから設定されている。このように設定された根拠となっている定理の名称を答えなさい。
- (4) 2チャンネルで記録した音楽 CD の 1 分の曲のデータ量を計算しなさい。ただし、音楽 CD は 44.1kHz で標準化し、標準化した振幅を 16 ビットで表現するものとする。なお 1MB は 1,024KB とし、MB に換算後の小数点以下は切り捨てる。
- (5) 20行目は動画ファイルを指定している。動画を構成する 1 枚の静止画像をフレームといい、1 秒間に画面に表示されるフレーム数をフレームレートという。フレームレートの単位として適切なものを、次のア～エの中から選び、記号で答えなさい。

ア bps

イ fps

ウ dpi

エ ppi

- (6) この HTML ファイルを改変し色の表現を学習する教材としたい。このため 5 行目と 6 行目の間に次の 1 行を追加した。このことによってページがどのように変化するかを答えなさい。

```
<style>body {background-color: #0000FF;}</style>
```

- (7) ディスプレイなどのみずから発光する画面の色は、光の三原色 (R, G, B) の組み合わせによってつくられる。これらの色の光を重ねると、もとの光よりも明るくなって白色に近づく。このような色のまざり方を何混色というか答えなさい。

検査Ⅲ 情 報

- 3 あるプログラムを用いて、「DeciNum」「iNum」「GouNum」を整数型、「DiviList」を文字型の変数と宣言し、下のようなプログラムA、プログラムBの順に実行し、「DiviList」及び「GouNum」の結果を表示するプログラムを作成した。次の各問いに答えなさい。

プログラム A

```
DeciNum に ① を代入する
GouNum に 0 を代入する
DiviList を 初期化する
iNum を 1 から ① まで 1 ずつ増やしながら繰り返す
    もし DeciNum Mod iNum が 0 に等しいとき
        GouNum に GouNum + iNum を代入する
        DiviList に DiviList & iNum & ", " を代入する
4行目に戻る
```

注1 「A Mod B」 AをBで割った余りを表す

注2 「, 」は「, (カンマ)」と「 (半角スペース)」の文字列を表す

プログラム B

```
もし Len(DiviList) が 0 より大きいとき
    DiviList に Left(DiviList, Len(DiviList) - 2) を代入する
```

注 「Len(A)」 変数Aに代入されている文字列の長さを取得する

- (1) プログラムAは、どんな処理を行っているのか30文字以内で答えなさい。
- (2) プログラムBは、どんな処理を行っているのか30文字以内で答えなさい。
- (3) ①を整数「187」にしてプログラムを実行した。このとき「DiviList」と「GouNum」の表示結果を答えなさい。
- (4) プログラムを実行したとき、「DiviList」の表示結果の一部が「1, 2, 4, 8, 16, 」で、「GouNum」の表示結果が整数「511」であった。このとき①の整数を答えなさい。
- (5) ①を整数「28」にしてプログラムを実行した。このとき「DiviList」の表示結果を答えなさい。

検査Ⅲ 情 報

- (6) 「DeciNum」に実数を代入して、このプログラムを実行したとき、「GouNum」及び「DiviList」の表示結果が下の表のようになった。「DeciNum」に実数を代入してプログラムを実行した際にどのような処理が行われると考えられるか答えなさい。

表

DeciNum	DiviList	GouNum
5.2	1, 5	6
5.3	1, 5	6
5.4	1, 5	6
5.5	1, 2, 3, 6	12
5.6	1, 2, 3, 6	12
5.7	1, 2, 3, 6	12

- (7) 「SquNum」と「kNum」の2つを整数型の変数として追加で宣言し、プログラム A を下のプログラム C に書き換え、①を整数「64」にしてプログラム C、プログラム B の順に実行した。このとき「DiviList」と「GouNum」の表示結果を答えなさい。

プログラム C

```

DeciNum に ① を代入する
SquNum に DeciNum の平方根を代入する
GouNum に 0 を代入する
DiviList を 初期化する
iNum を 1 から SquNum まで 1 ずつ増やしながら繰り返す
  もし DeciNum Mod iNum が 0 に等しいとき
    kNum に DeciNum / iNum を代入する
    もし kNum が iNum に等しいとき
      GouNum に GouNum + iNum を代入する
      DiviList に DiviList & iNum & ", " を代入する
    それ以外るとき
      GouNum に GouNum + iNum + kNum を代入する
      DiviList に DiviList & iNum & ", " & kNum & ", " を代入する
5行目に戻る

```

検 査 Ⅲ 情 報

4 次の表は、ある家庭内 LAN にある端末の設定情報である。次の各問いに答えなさい。
表

行	設定情報
1	物理アドレス : FF-7A-91-3E-F6-91
2	DHCP 有効 : はい
3	IPv6 アドレス : 240b:12:6621:b300:2da4:a903:ef38:c2f9
4	IPv4 アドレス : 192.168.11.29
5	サブネット マスク : 255.255.255.0
6	デフォルト ゲートウェイ : 192.168.11.1
7	DHCP サーバー : 192.168.11.1
8	DNS サーバー : 192.168.11.1

(1) 次の先生と生徒の会話文の①～⑩にあてはまる適切な語句、または数値を答えなさい。

[先生と生徒の会話文]

先生：この情報は、ある家庭内 LAN の端末の設定情報です。これを見ていきましょう。1 行目の物理アドレスは 1 バイトずつ 6 個のブロックに分け、これを先頭から 16 進数 2 桁ずつの数字にし、ハイフン (-) で繋ぎ表されていますが、(①) ビットの 2 進数のアドレスです。ROM に焼き付けられていて簡単に書き換えられません。アルファベット 3 文字を使って (②) アドレスとも呼ばれます。

生徒：そのほかにも何々アドレスという単語があります。これらは何ですか？

先生：IP アドレスは現在、IPv4 アドレスと IPv6 アドレスが使われています。バージョンの新しい IPv6 は 128 ビットで構成され、これを 16 ビットごとに「:」で区切り、1 つのセクションを (③) 進数 4 桁で表記するのが基本となっています。

生徒：なぜ IPv4 と IPv6 という二つのバージョンがあるのですか？

先生：それは IP アドレス (IPv4) の (④) という問題が発生することが予想され、たくさんの IP アドレスが使えるバージョンとして IPv6 が利用され始めたからです。ちなみに IPv4 では、原理的には 2^{32} 個、およそ (⑤) 億弱の IP アドレスが設定できますが、インターネットの利用者の増加や端末台数の増加によって先ほどの問題が心配されています。

生徒：IPv6 アドレスを使う他には対策がないのですか？

先生：IPv4 アドレスを節約することを目的に、IP アドレスを (⑥) IP アドレスとプライベート IP アドレスに分ける方法があります。(⑥) IP アドレスは世界中で同じ IP アドレスが存在してはいけません。プライベート IP アドレスは、隣の家で同じ IP アドレスを使用することができます。

生徒：家庭内 LAN からインターネットへ接続するにはどうするのですか？

先生：このとき利用される機材が (⑦) です。ここで LAN 側の端末のプライベート IP アドレスを (⑥) IP アドレスに変換します。表から (⑦) の LAN 側の IP アドレスは、(⑧) であることがわかります。

検 査 Ⅲ 情 報

IPv4 でインターネット接続する PC にネットワーク設定をする場合は、機器に IP アドレス、(⑨), デフォルトゲートウェイを設定する必要があります。

生徒：8行目にあるDNSというのは何ですか。

先生：DNS を元の英単語にすると (⑩) Name System となります。これはインターネット上のサーバーのドメイン名を IP アドレスに変換する仕組みです。数字で表されている IP アドレスを人間が扱いやすい文字列で示したドメイン名に変換します。

生徒：この家庭内 LAN では、6行目のデフォルトゲートウェイ、7行目のDHCPサーバー、8行目のDNSサーバーが同じIPv4アドレスです。問題ないのですか？

先生：問題ありません。Wi-Fi ルーターが1台で、これらの機能を担っているでしょう。

(2) 以下の〔説明〕はクラウドサービスのサービスモデルについて述べている。説明とサービスモデルの名称の適切な組み合わせを、下の〔組み合わせ〕ア～カの中から選び、記号で答えなさい。

〔説明〕

- ⅰ ネットワーク経由で、共有ディスクや OS を自由にインストール可能な仮想マシンなどハードウェアやインフラ機能を提供する。
- ⅱ ネットワーク経由で、仮想化されたアプリケーションサーバーやデータベースなどアプリケーション実行用のプラットフォーム機能を提供する。
- ⅲ ネットワーク経由で、電子メール、グループウェア、顧客管理、財務会計などのソフトウェアを提供する。

〔組み合わせ〕

ア	ⅰ	SaaS	ⅱ	PaaS	ⅲ	IaaS	イ	ⅰ	SaaS	ⅱ	IaaS	ⅲ	PaaS
ウ	ⅰ	PaaS	ⅱ	IaaS	ⅲ	SaaS	エ	ⅰ	PaaS	ⅱ	SaaS	ⅲ	IaaS
オ	ⅰ	IaaS	ⅱ	PaaS	ⅲ	SaaS	カ	ⅰ	IaaS	ⅱ	SaaS	ⅲ	PaaS

検査Ⅲ 情 報

- 5 ある会社では、社内の備品を有効活用するため、関係データベースを利用して貸出管理を行っている。次の各問いに答えなさい。

[貸出管理の流れ]

- i 備品の利用希望者は、図1の備品貸出申込書に必要事項を記入して提出する。
- ii 備品管理担当者は、備品貸出申込書の記入内容を表3の貸出表4月に入力する。なお、貸出時には「返却日」に0を入力する。
- iii 備品の返却時に、表3の貸出表4月「返却日」を更新する。

図1

備品貸出申込書	
部署コード	S101
部署名	経理課
利用者名	土屋三郎
備品コード	K2012
備品名	ノート型パソコン
貸出日	2025年4月10日
返却予定日	2025年4月30日
利用目的	展示会でのプロモーション

〇〇〇〇株式会社

表1 部署表

部署コード	部署名
S101	経理課
S102	人事課
S103	総務課
S201	営業1課
S202	営業2課
S203	販売促進課
S301	研究開発課
S302	広報課

表2 備品表

備品コード	備品名
K1001	ホワイトボード
K1002	拡声器
K2011	デスクトップ型パソコン
K2012	ノート型パソコン
K2013	タブレットパソコン
K3011	デジタルカメラ
K3012	液晶プロジェクタ
K3013	スクリーン

表3 貸出表4月

貸出番号	部署コード	利用者名	備品コード	貸出日	返却予定日	返却日
254051	S203	香山一郎	K2012	20250403	20250405	20250405
254052	S202	川治次郎	K1001	20250403	20250405	20250405
254087	S101	北林七郎	K2012	20250405	20250418	0
254097	S302	萩田八郎	K2013	20250408	20250420	0
254110	S302	富本十郎	K3011	20250409	20250422	0
254193	S101	土屋三郎	K2012	20250410	20250430	0
254194	S202	古河史郎	K3012	20250410	20250430	0
254195	S202	古河史郎	K3013	20250410	20250430	0
254196	S302	中西五郎	K3011	20250420	20250423	20250423
}	}	}	}	}	}	}

※表3の貸出表4月「貸出日」、「返却予定日」、「返却日」は、8桁の数値を用いる。ただし、「返却日」に0が入力されている場合は、未返却を表す。なお、20250401は2025年4月1日を表す。

- (1) 次のSQL文を実行する場面の説明として適切なものを、下のア～ウの中から選び、記号で答えなさい。

UPDATE 貸出表4月 SET 返却日 = 20250425 WHERE 貸出番号 = 254193

- ア 2025年4月25日に貸出番号が254193の備品の返却を受けた際の処理
- イ 貸出番号が254193の備品の返却予定日を2025年4月25日変更するための処理
- ウ 貸出表4月から貸出番号が254193の備品を削除するための処理

検査Ⅲ 情 報

- (2) 次の SQL 文を実行して、返却予定日を過ぎている貸出中の備品について、部署名、利用者名、備品名、超過日数を調べる表4を作成したい。①は空欄にあてはまる適切なフィールド名を答えなさい。②は空欄にあてはまる適切なものを、下のア～ウの中から選び、記号で答えなさい。ただし、部署コードの降順、同一部署の中では貸出日の昇順に並べ替える。なお、本日は2025年4月25日である。

表4 返却予定日超過者一覧表

部署名	利用者名	備品名	超過日数
広報課	萩田八郎	タブレットパソコン	5
広報課	富本十郎	デジタルカメラ	3
経理課	北林七郎	ノート型パソコン	7

```
SELECT 部署名, 利用者名, 備品名, 20250425 - 返却予定日 AS ( ① )
FROM 部署表, 備品表, 貸出表 4 月
WHERE 部署表. 部署コード = 貸出表 4 月. 部署コード
      AND 備品表. 備品コード = 貸出表 4 月. 備品コード
      AND 返却予定日 < 20250425
      AND 返却日 = 0
ORDER BY 部署表. 部署コード ( ② )
```

ア DESC, 貸出日 ASC イ ASC, 貸出日 DESC ウ DESC, 貸出日 DESC

- (3) 次の SQL 文を実行して、表3と同様の形式で作成された貸出表3月から、備品ごとの貸出回数を調べる表5を作成する場合、①は空欄にあてはまる適切な関数を答えなさい。②は空欄にあてはまる適切なものを、下のア～ウの中から選び、記号で答えなさい。なお、貸出表3月は貸出日が2025年3月1日から2025年3月31日までの内容である。

表5 貸出回数一覧表

備品コード	備品名	貸出回数
K1001	ホワイトボード	13
K2012	ノート型パソコン	9
K2013	タブレットパソコン	5
K3011	デジタルカメラ	14
K3012	液晶プロジェクタ	2

```
SELECT 備品表. 備品コード, 備品名, ( ① ) AS 貸出回数
FROM 備品表, 貸出表 3 月
WHERE 備品表. 備品コード = 貸出表 3 月. 備品コード
      AND ( ② )
GROUP BY 備品表. 備品コード, 備品名
```

ア 貸出日 > 20250301 AND 貸出日 < 20250331
 イ 貸出日 BETWEEN 20250301 AND 20250331
 ウ 貸出日 <= 20250301 AND 貸出日 >= 20250331

- (4) 営業1課と営業2課を統合するため、部署表から営業2課の行を削除したい。しかし、この行(部署コード S202)は貸出表4月と関係をもっており、削除しようとするエラーが生じた。その理由として適切なものを、次のア～ウの中から選び、記号で答えなさい。

ア 排他制御が起きるため
 イ 整合性制約(参照整合性)に反するため
 ウ デッドロックが発生するため

検査Ⅲ 情報

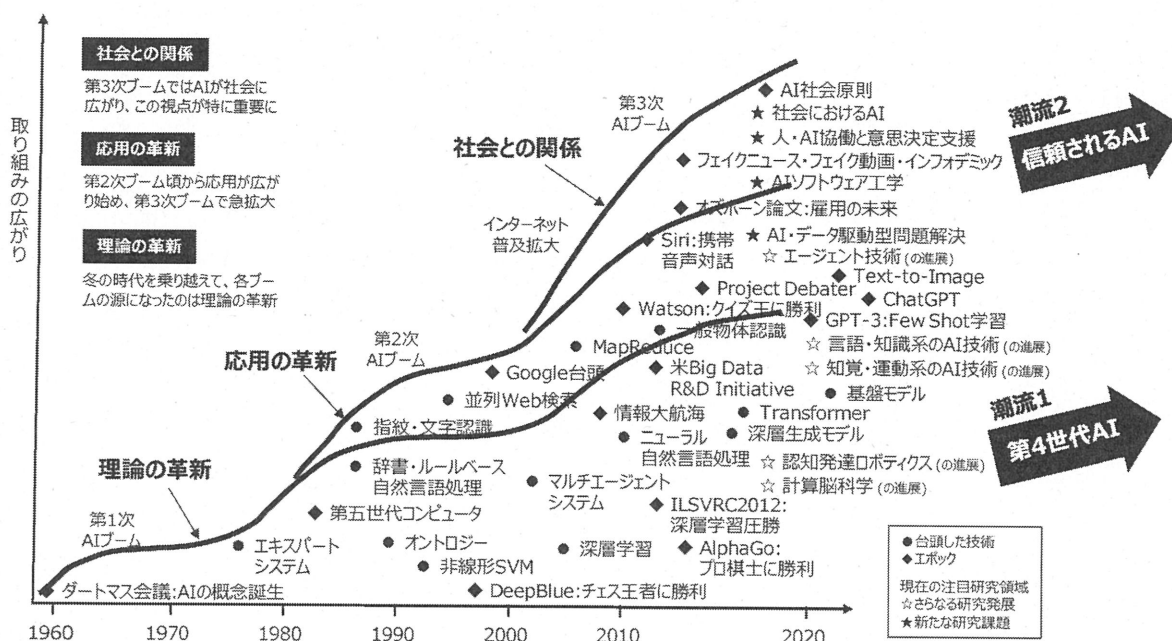
6 次の文章を読んで、各問いに答えなさい。

AI 技術は、図のように、幾度もブームを繰り返しながら、研究開発の取組が広がってきています。

図 人工知能・ビッグデータ技術の俯瞰図（時系列）

科学技術振興機構研究開発戦略センター（2023）

「人工知能研究の新潮流2～基盤モデル・生成AIのインパクト～」より引用



1950年代後半から1960年代に起きた「第1次ブーム」では、AIに関する基礎的な概念が提案され、新しい学問分野として立ち上がりました。1966年には、(A) 初の対話型自然言語処理プログラムが開発されました。

1980年代から1990年代に起きた「第2次ブーム」では、人手で辞書・ルールを構築・活用する手法が主流となりましたが、人手による限界などから、一度ブームは終息しました。

そして、2000年代以降の技術の進展に伴うインターネットや計算能力の拡大を背景として、大量のデータから帰納的にルールやモデルを構築する（①）技術の進展などによって、2010年頃から「第3次ブーム」が牽引（けんいん）されてきましたが、さらに最近の生成AI技術の進展などを受けて、現在は「第4次ブーム」に差し掛かっているとも言われています。

（①）技術には、様々な「モデル」と「学習法」がありますが、特に近年発展しているモデルは、人間の脳の神経細胞（ニューロン）が回路網を形成して情報伝達している仕組みを参考に作られた、人工ニューロンのネットワーク（ニューラルネットワーク）を多層化・大規模化して用いる「（②）」です。このニューラルネットワークのモデルは1960年前後に活発に研究され、近年になって、計算機の性能向上や利用可能なデータ量の増大とあいまって、急速に技術の発展が進みました。

（②）の研究開発は、グラフィックスの描画に用いられていたプロセッサを計算に用

検査Ⅲ 情報

いることで大きく進展しました。

2020 年前後から関心の高まっている画像生成や対話型の AI などの「生成 AI」では「自己教師あり学習」が用いられています。このうち、近年、「大規模言語モデル」と呼ばれる、言語を中心とした（ ② ）モデルの研究が飛躍的に進みました。中でも、自然言語で指示や条件（プロンプト）を与えることができる「対話型生成 AI」は、一般の人々の仕事や生活での利用も急速に広がりました。

AI は、その研究や開発の過程において、適用範囲や能力により、大きく二つのカテゴリに分けることができます。特定の機能や特定の状況下でのみ人間に近く、時には精度で人間を上回る動作をする「特化型 AI」と、人間の知能の様々な側面を幅広くカバーし、様々な状況で人間の知能のように動作する「（ ③ ）」です。第 3 次ブームにおいて、様々な応用に広がった AI 技術は基本的に「特化型 AI」に相当する技術群とされていますが、極めて大規模な（ ② ）によって作られた「基盤モデル」の登場により、研究開発の中心がこれまでの特化型 AI と言われていたものから、高いマルチモーダル性・汎用性を持つ AI へと向かいつつあります。

このように、生成 AI に関する技術の進展や、技術者や専門家でなくてもインターネット上で自然言語を用いて利用できるようになってきていることなどにより、生成 AI 技術を活用した市場は今後も拡大すると見込まれています。

- (1) 下線部 (A) のプログラムの名称を、次のア～エの中から選び、記号で答えなさい。
ア ENIGMA イ ENIAC ウ ELIZA エ EDSAC
- (2) ①～③に入る適切な語句を、それぞれ次のア～キの中から選び、記号で答えなさい。
[語群]
ア LLM イ 機械学習 ウ 生成 AI エ 深層学習 オ 汎用 AI
カ GPT キ ニューロン
- (3) 機械学習の学習法としては主に「教師あり学習」、「教師なし学習」、「強化学習」があります。これらの学習法の説明として適切なものを、それぞれ次のア～ウの中から選び、記号で答えなさい。
ア 一定の環境の中で試行錯誤を行い、報酬を与えることにより学習するモデル
イ 正解のラベルを付けない学習データにより学習するモデル
ウ 正解のラベルを付けた学習データで学習するモデル

検査Ⅲ 情報

7 次の文章を読んで、各問いに答えなさい。

一般に○（マル）と×（バツ）のような2種類のデータを分けるような直線を決定境界といい、図1及び図2のように決定境界は無数に存在する。

図1

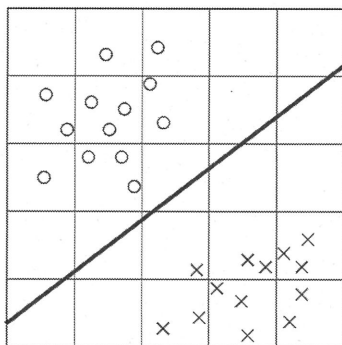
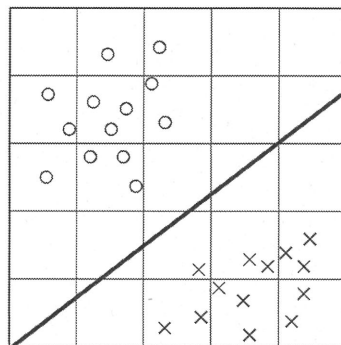


図2



決定境界とデータの間の距離のうち、最も短いものをマージン (margin) と呼ぶ。

図3の直線の例では、直線がマルの近くに引かれている。どのバツのデータとも大きく離れていることになるが、マージンは最も近いデータとの距離であるため、この直線とデータのマージンはかなり小さいことがわかる。

一方で、図4を見てみると、直線はマルのデータとも、バツのデータとも、適度に離れていることがわかる。このように、マージンが最大になる決定境界を求めることが重要である。

図3

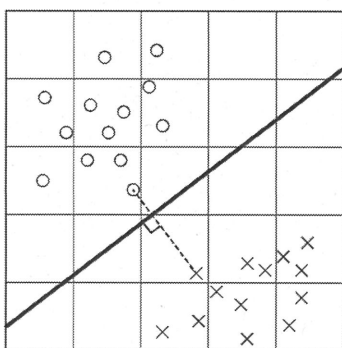
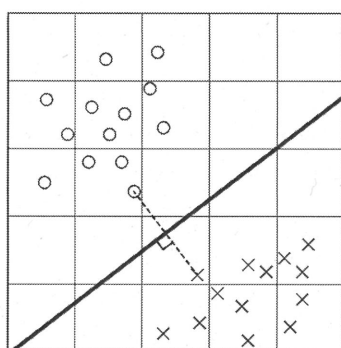


図4



(1) 次の文章は、機械学習の1つを説明したものである。この機械学習の名称を答えなさい。

「教師あり学習」によるパターン認識モデルで、データの分類や回帰分析などで利用されている。訓練データをグループに分ける決定境界を利用した分類が行われるが、新たなデータの分類精度が高い傾向にある。

検査Ⅲ 情報

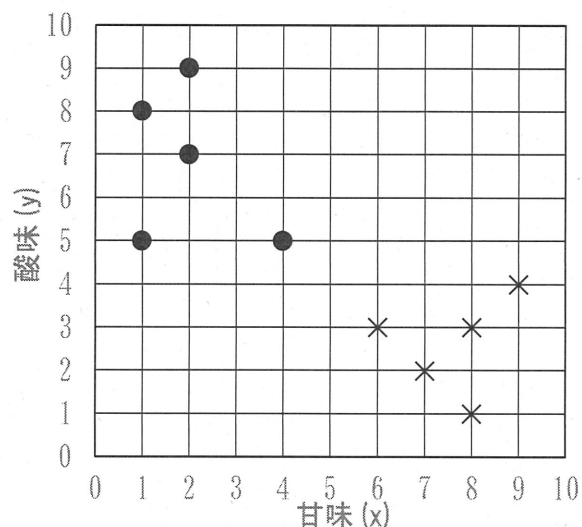
- (2) 次の表は、レモンとメロンそれぞれ5個ずつについて甘味と酸味をそれぞれ10段階でラベル付けをしたデータであり、図5はそれらをプロットしたものである。

図5において、レモンとメロンを分類するために、マージンが最も大きくなる決定境界を次の手順に従って求め、直線 $y=ax+b$ の形式で答えなさい。

表

甘味 (x)	酸味 (y)	ラベル
1	5	レモン
1	8	レモン
2	9	レモン
2	7	レモン
4	5	レモン
6	3	メロン
8	1	メロン
7	2	メロン
8	3	メロン
9	4	メロン

図5



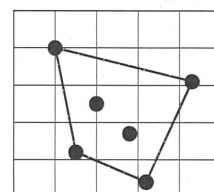
手順

- (i) それぞれのデータの凸包を求める。
- (ii) 凸包上の一番近い2点を見つける。
- (iii) その2点を結ぶ線分の垂直二等分線を求める。

凸包とは

図6は、6つのデータを平面上にプロットしたものである。この6つのデータの場合に釘が刺さっている際に、輪ゴムを外側から大きく広げた後に手を放すと図6のように引っかかるときにできる多角形のことを凸包と呼ぶ。

図6



- (3) (2) で求めた決定境界を用いると、甘味が3、酸味が3の場合は、レモンとメロンのどちらに分類されるか。その理由も答えなさい。