

AI 基礎プログラミング

第1回 効果測定 問題

問題1 配点 24 点

AI・機械学習に関する以下の説明に当てはまる用語を解答群から選択して記号を答えなさい。

AIは大きく、【①】と【②】の2種類に分けられる。人間のように1人でさまざまなことに対処できるAIを【①】と呼ぶ。一方、特定の領域でのみ、人間と同等以上の判断を行うAIを【②】と言う。なお、AIには明確な定義が【③】。

機械学習の手法の1つである、【④】では、入力データと答えのデータのペアとなる【⑤】をコンピュータに大量に渡し、データ間の法則を探させる。このデータ間の法則のことを【⑥】と呼び、この法則を明確にする過程を学習と呼ぶ。

【④】では、学習で使用したデータでは予測性能が高いに関わらず、学習に使用していないデータでの予測性能が低くなるという現象が起こることがある。この状態のことを【⑦】と呼ぶ。また、この状態になるのを避けつつ予測性能を向上させるため、一般的にデータを【⑧】、【⑨】、【⑩】に分割し、学習と評価を行う。

解答群

ア 教師あり学習	イ 教師なし学習	ウ ホールドアウト法
エ 過学習	オ 強化学習	カ 超過学習
キ ディープラーニング	ク 弱いAI(特化型 AI)	ケ 強い AI(汎用型 AI)
コ 教師データ	サ テストデータ	シ 検証データ
ス 正解データ	セ 訓練データ	ソ 補足データ
タ ある	チ ない	ツ モデル

AI 基礎プログラミング

第1回 効果測定 問題

問題2 配点6点

機械学習によるデータ分析の流れになるように並べたとき、以下の①～⑥に当てはまる内容を解答群から選択して記号を答えなさい。 ※ア～オの全てを使用すること。また、同じ記号を使用してもよい。

1. 上司からデータ分析について相談があった
2. ____①____
3. 分析で使用するため、インターネットで公開されているデータを収集した
4. ____②____
5. ____③____
6. ____④____
7. ____⑤____
8. ____⑥____
9. 完成したモデルを使用し、データ分析を実施した

解答群

- | | |
|---|----------------------------------|
| ア | モデルにデータを与えて学習させた |
| イ | 複数のデータを結合した |
| ウ | 課題を明らかにし、機械学習を用いることに決定した |
| エ | モデルの精度が目標に至ったことを確認した |
| オ | モデルの精度が目標に至らないため、特徴量に新しいデータを追加した |

AI 基礎プログラミング

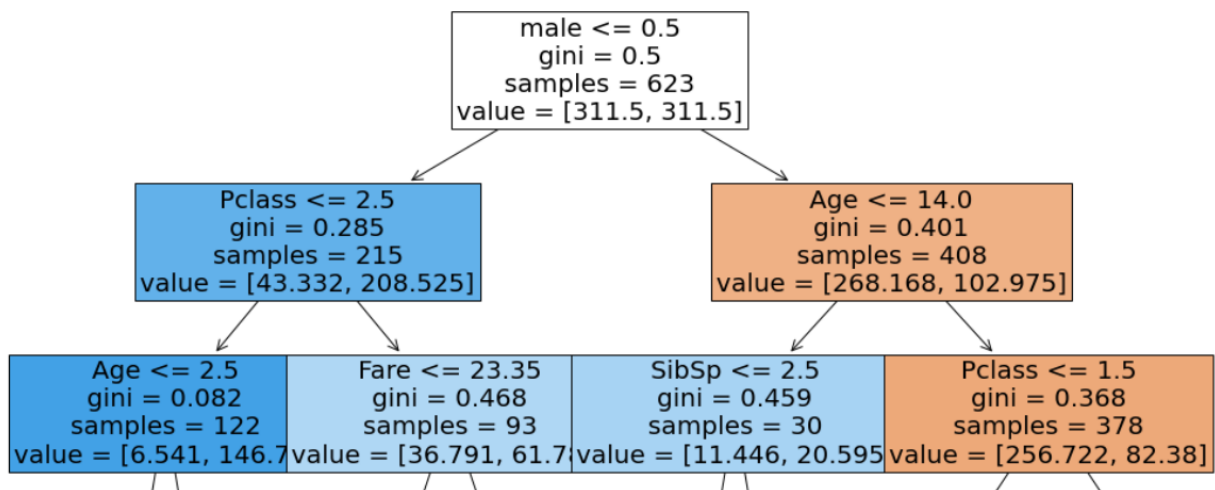
第1回 効果測定 問題

問題3 配点 40 点

配布データ「Survived_kouka1.csv」を用い、「Survived」列のデータを予測するプログラムを作成する。

- (1) 下記の実行結果となるように、未完成リストのコメントに従って実装し解答欄に記入してください。
※穴埋め箇所以外の実装済みプログラムを修正しないこと
- ① 「Survived」列のデータに偏りが有るかどうか(データごとの出現回数)を確認する。
 - ② 「Survived」列を基準に、「Age」列の平均値を表示する。
 - ③ 「Survived」列のデータと「Pclass」列でクロス集計した中央値を表示する。
※縦軸を「Survived」列、横軸を「Pclass」列とすること。
 - ④ 「Age」列で、条件に当てはまる箇所の値を書き換える。(条件はプログラム内のコメント参照)
 - ⑤ 「Sex」列をダミー変数化する。
 - ⑥ 特徴量を格納したデータフレームと上記ダミー変数化したものを横方向に連結する。
 - ⑦ データを「学習で用いるデータ」と「テストで用いるデータ」に分割する。
 - ⑧ 不均衡データを考慮した決定木を生成する。
 - ⑨ 学習済みの決定木で、「学習で用いるデータ」と「テストで用いるデータ」の正解率を表示する。
 - ⑩ 学習済みの決定木で、分岐条件に使用している列、分岐条件の閾値を表示する。
 - ⑪ 学習済みの決定木を保存する。

- (2) 決定木の分岐条件を図示すると3階層までは下記であった。このとき下記の問いに答えなさい。



- ① 初めの分岐条件として「male」列の閾値(下記条件式の空欄)を答えなさい。
male ≤ 空欄
- ② 「male」列が0だった場合、次にどの列を条件に分岐するか、下記から選択して記号を答えなさい。

解答群

ア male	イ Pclass	ウ Age	エ Fare	オ SibSp
--------	----------	-------	--------	---------

AI 基礎プログラミング

第1回 効果測定 問題

問題4 配点 30 点

配布データ「Boston_kouka1.csv」を用いて、「価格」列のデータを予測するプログラムを作成する

- (1) 下記の実行結果となるように、未完成リストのコメントに従って実装し解答欄に記入してください。
- ※穴埋め箇所以外の実装済みプログラムを修正しないこと
- ① 横軸を「部屋数」、縦軸を「価格」として散布図を作成する。
 - ② 上記の散布図で一番右下の外れ値を削除する。
 - ③ 「価格」とその他の列の相関係数を表示する。（「価格」同士の相関係数を含める）
 - ④ 上記の相関係数の絶対値を降順に表示する。
 - ⑤ 「部屋数」の2乗を「部屋数2」という名前で特徴量のデータフレームに追加する。
 - ⑥ 「学習で用いるデータ」を標準化する。
 - ⑦ 回帰を使って学習させる。
 - ⑧ 「テストで用いるデータ」を使用して、決定係数を表示する。
 - ⑨ 各カラムの係数、切片を確認する。
 - ⑩ 未知のデータでの予測結果を求める（データの内容はプログラム内のコメント参照）。