

機械学習テスト練習問題

Copyright 2021 DataMix Co., Ltd.

本練習問題の対象者

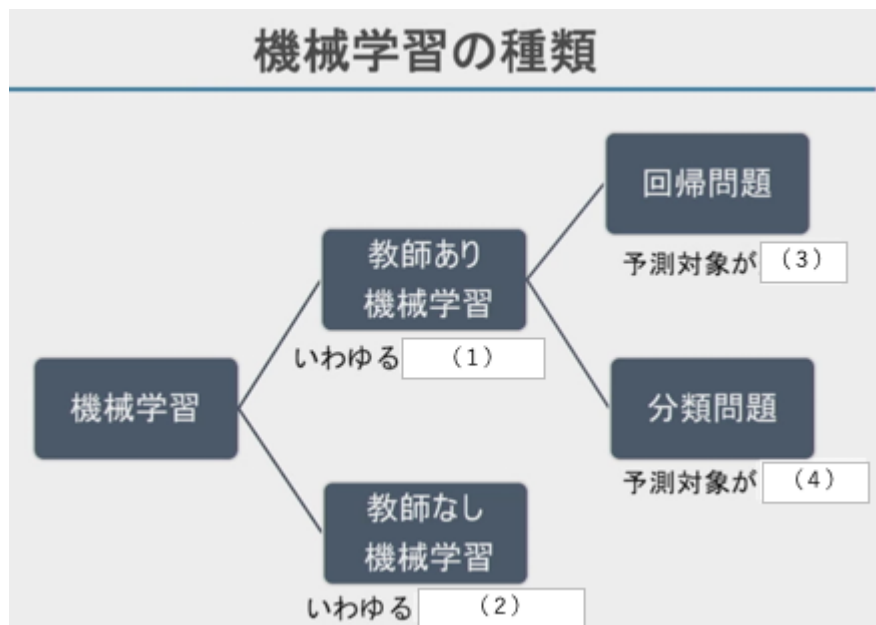
機械学習に関する一定の知識があることを前提に作られています。

本試験に関する注意事項

- こちらの練習問題は、入学試験で出題される問題の類題となっております。
- 実際の試験は、HackerRank (<https://www.hackerrank.com>) というサービスで実施されます。
- HackerRankのサイトは英語ですが、試験は日本語です。

Q1. 機械学習の種類

下図の空欄を埋めよ。



選択肢

a, 仮説検定、b, 予測モデル、c, グルーピング、d, 時系列、e, 連続値、f, カテゴリ
g, 整数

Q2. Pythonを使ったデータ読み込み

Python で 'data.csv' というcsvファイルをPandasのDataFrame型で読み込み dataset という変数に代入するコードを選べ。

a, dataset <- read_csv('data.csv')
b, dataset = read.csv('data.csv')
c, dataset = pd.read_csv('data.csv')
d, dataset.to_csv('data.csv')

Q3. 線形回帰

線形回帰に関する以下の文章の空欄に当てはまる適切な選択肢を1つ選択しなさい。

線形回帰分析とはパラメータの(1)と(2)を調整することで、目的関数である(3)を(4)することである。

選択肢

a,分散、b,切片、c,残差の二乗の合計、d,ジニ係数、e,傾き、f,平均、g,最小化、h,最大化

Q4. 勾配降下法

$f(x) = x^2 - 3x + 3$ という関数に対して、勾配降下法を適用して最小値を見つけようとしています。

初期値 $x_1 = 3$ 、学習率を0.1としたときの次の x_2 を求めよ。

- a, 3.3
- b, 2.7
- c, 3.03
- d, 3.5

Q5. ジニ係数

10人中3人が解約、7人が継続した場合のジニ係数を計算せよ。

- a, 0.67
- b, 0.58
- c, 0.33
- d, 0.42

Q6. ハイパーパラメータチューニング

ハイパーパラメータチューニングに関する文として正しいものを全て選択しなさい。

a, クロスバリデーションは、データを複数のサブセットに分割し、ある1つのサブセットを検証用、その他のサブセットを学習に用いてモデルを作り精度を検証し、サブセットの数だけ検証を繰り返し行う。

b, ランダムサーチとは、ランダムフォレストのように複数のハイパーパラメーターがある場合、候補として挙げたハイパーパラメーターの全ての組み合わせをクロスバリデーションで検証し、最も精度の高いハイパーパラメーターの組み合わせを探索する手法である。

c, クロスバリデーションは、未知のデータに対する精度(汎化性能)を推定するための手法である。

d, グリッドサーチとは、ランダムフォレストのように複数のハイパーパラメーターがある場合、候補として挙げたハイパーパラメーターの全ての組み合わせをクロスバリデーションで検証し、最も精度の高いハイパーパラメーターの組み合わせを探索する手法である。

Q7. ランダムフォレスト

以下の空欄①～④に当てはまる適切な言葉の組み合わせを一つ選びなさい。

ランダムフォレストとは、【①】の一種で、多くの決定木を作り【②】によって最終的な予測値を得る教師あり学習のアルゴリズムである。ひとつの決定木を作成する際には、【③】することと同時に、列もサンプリングすることで決定木の多様化に成功し、精度が高くなりやすい手法である。ランダムフォレストの主なハイパーパラメーターは木の数、最大の木の深さ、【④】である。

a, ① アンサンブル学習 ② 多数決 ③ 重複を許さずサンプリング ④ 多数決に使用する決定木の数

b, ① 教師なし学習 ② 最も精度の高い一つの決定木 ③ 重複を許してサンプリング ④ 使用する列の個数

c, ① アンサンブル学習 ② 最も精度の高い一つの決定木 ③ 重複を許さずサンプリング ④ 多数決に使用する決定木の数

d, ① アンサンブル学習 ② 多数決 ③ 重複を許してサンプリング ④ 使用する列の個数

Q8. K-Means法

以下の空欄①～②に当てはまる適切な言葉の組み合わせを一つ選びなさい。

K-Means法のアルゴリズム

- 1、はじめにk個の【①】を置く
- 2、各サンプルを最も近い【①】と同じ【②】に分類する
- 3、k個の【②】それぞれで重心を求めそれを新たな【①】として更新する
- 4、重心の位置が変化しなくなるまで2～3を繰り返す

選択肢

a,センター b,クラスター c,シード d,フラグ

Q9. K-Means法 最適なクラスター数

K-means法をいくつかのクラスター数で実行した結果以下の表のような結果が得られた。

クラスター数	1	2	3	4	5	6
目的関数の値	11.2	7.7	4.5	2.1	1.9	1.8

エルボー法による最適なクラスター数はいくつか答えなさい。

- 1
- 2
- 3
- 4
- 5
- 6

Q10. 次元圧縮

以下の中から次元圧縮の手法を全て選べ

- a, 特異値分解
- b, 線形回帰分析
- c, ロジスティック回帰分析
- d, 非負値行列分解
- e, 主成分分析
- f, クラスタリング