

## 情報工学実験 2 10/25 課題

### 1 実験内容

scikit-learn ライブラリを用いて, 様々な分類モデルを作成し, 性能評価を行った.

#### データセット

用いたデータセットは scikit-learn ライブラリに組み込まれている Iris データセットを用いた. 「花びらの長さ」, 「花びらの幅」の 2 つの特徴量を用いて, Iris-setosa, Iris-versicolor, Iris-virginica の 3 つのラベルを 0, 1, 2 と整数で符号化したクラスラベルに対して分類を行った.

全体の 30% をテストデータ, その他 70% を学習データとしてデータセットをランダムに分割し実験を行った. データを分割する際, クラスラベルの比率は入力データセットと等しくなるようにした.

#### 用いたアルゴリズム

分類モデルを作成する際, 以下のアルゴリズムを用いた.

- パーセプトロン
- ロジスティック回帰
- 線形 SVM
- カーネル SVM
- 決定木
- ランダムフォレスト

#### 性能評価

モデルの性能を評価する際, テストデータセットを用いて, 以下の指標で評価した.

- 混同行列 (Confusion matrix)
- 正解率 (Accuracy)
- 適合率 (Precision)
- 再現率 (Recall)
- F1 値 (F1-measure)

適合率, 再現率, F1 値に関しては各クラスごとと, 分類結果全体の計算を行った. なお分類結果全体の計算は, Iris データセットにおいてラベルごとのデータの偏りは起こっていないためマクロ平均で行った.[1]

### 2 実験結果

各アルゴリズムを用いたモデルに対し, 前節で述べた性能評価を行った結果を示す.

#### パーセプトロン

図 1 にパーセプトロンにおける混同行列を示す.

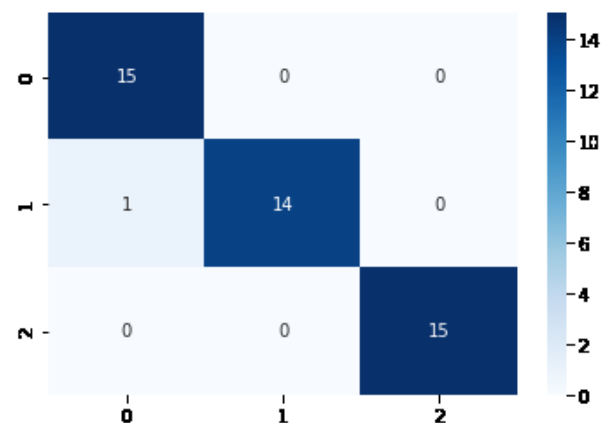


図 1: パーセプトロン分類モデルにおける混同行列

正解クラスが 0 であるデータに対してクラス 1 と誤分類したデータが 1 つあることがわかる.

また, 表 1 にパーセプトロンモデルにおける正解率, 適合率, 再現率, F1 値を示す.

表 1: パーセプトロンモデルにおける正解率, 適合率, 再現率, F1 値

	Precision	Recall	F1-measure
クラス 0	0.93750	1.00000	0.96774
クラス 1	1.00000	0.93333	0.96552
クラス 2	1.00000	1.00000	1.00000
分類結果全体	0.97917	0.97778	0.97775
Accuracy	0.97778		

ロジスティック回帰

線形 SNM

カーネル SVM

決定木

ランダムフォレスト

## 参考文献

- [1] 静かなる名辞. 【python】分類タスクの評価指標の解説と sklearn での計算方法, 2018-03-14. <https://www.haya-programming.com/entry/2018/03/14/112454>.