
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, 性能評価をした.

1.1 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した.

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した. 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.

1.2 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
- k 最近傍法

1.3 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.

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

適合率, 再現率, F1 値に関しては各クラスごと, 分類結果全体の計算をした. なお分類結果全体は, 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計算した.[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した結果を示す. 以下, 各モデルの正解率, 適合率, 再現率, F1 値を示した表と, 混同行列のヒートマップを示す.

2.1 パーセプトロン

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

	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		

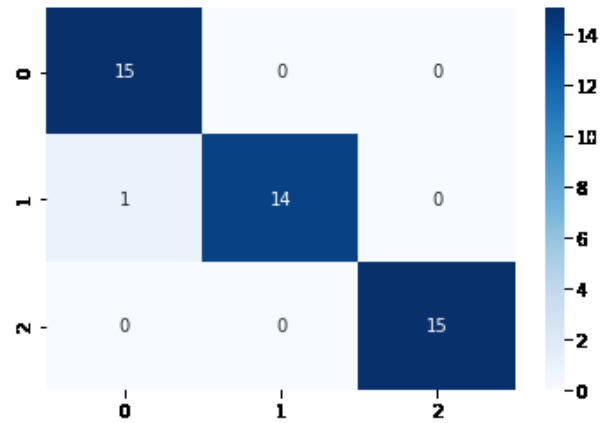


図 1: パーセプトロンモデルにおける混同行列のヒートマップ

2.2 ロジスティック回帰

表 2: ロジスティック回帰モデルにおける適合率, 再現率, F1 値, 正解率

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

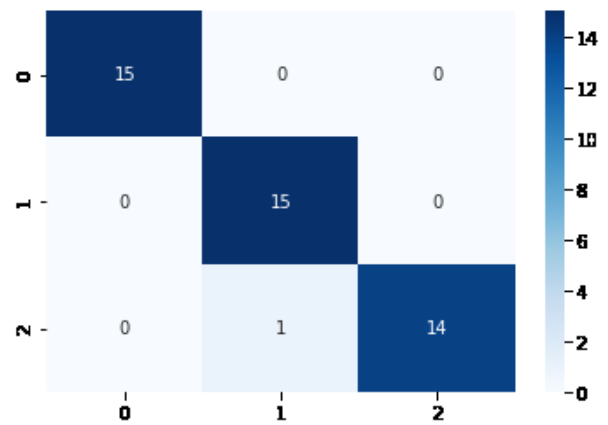


図 2: ロジスティック回帰モデルにおける混同行列のヒートマップ

2.3 線形 SVM

表 3: 線形 SVM 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	1.00000	1.00000	1.00000
クラス 1	0.93750	1.00000	0.96774
クラス 2	1.00000	0.93333	0.96552
分類結果全体	0.97917	0.97778	0.97775
Accuracy	0.97778		

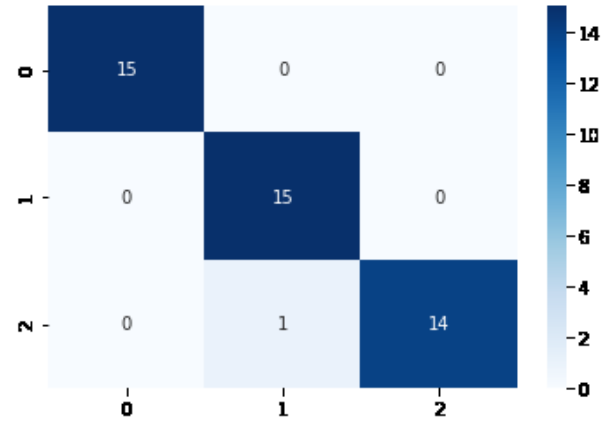


図 3: 線形 SVM モデルにおける混同行列のヒートマップ

2.4 カーネル SVM

2.4.1 パラメータ $\gamma = 0.2$ の時

表 4: カーネル SVM モデルにおける正解率, 適合率, 再現率, F1 値 ($\gamma = 0.2$)

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

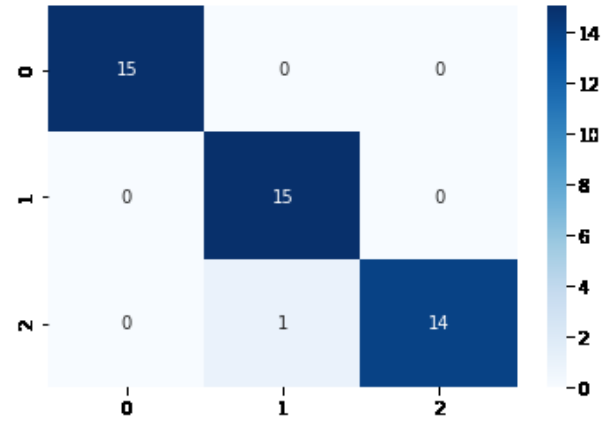


図 4: カーネル SVM モデルにおける混同行列のヒートマップ ($\gamma = 0.2$)

2.4.2 パラメータ $\gamma = 100.0$ の時

表 5: カーネル SVM モデルにおける正解率, 適合率, 再現率, F1 値 ($\gamma = 100.0$)

	Precision	Recall	F1-measure
クラス 0	1.00000	0.93333	0.96552
クラス 1	0.93333	0.93333	0.93333
クラス 2	0.87500	0.93333	0.90323
分類結果全体	0.93611	0.93333	0.93403
Accuracy	0.93333		

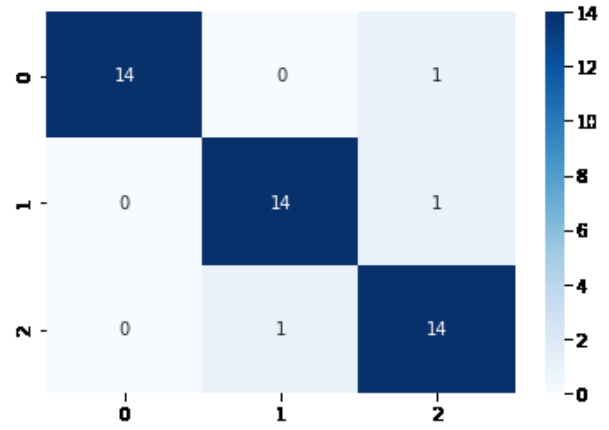


図 5: カーネル SVM モデルにおける混同行列のヒートマップ ($\gamma = 100.0$)

2.5 決定木

表 6: 決定木モデルにおける正解率, 適合率, 再現率, F1 値 ($\text{maxdepth} = 4$)

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

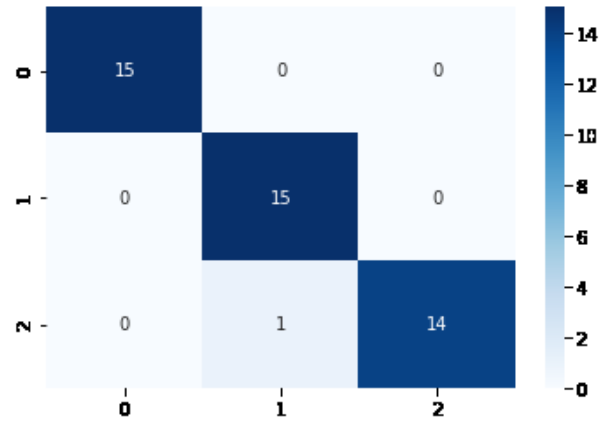


図 6: 決定木モデルにおける混同行列のヒートマップ ($\text{maxdepth} = 4$)

2.6 ランダムフォレスト

表 7: ランダムフォレストモデルにおける正解率, 適合率, 再現率, F1 値 (n_estimators = 25)

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

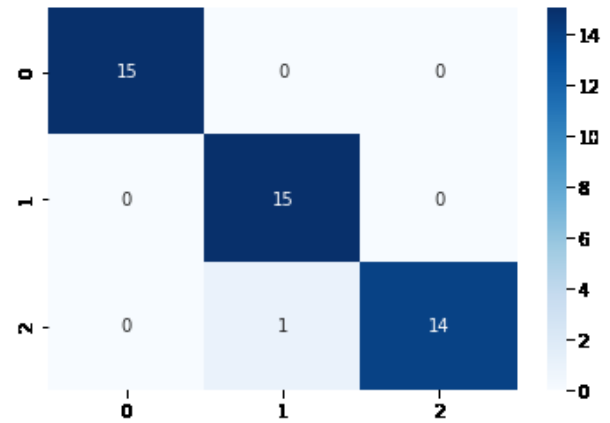


図 7: ランダムフォレストモデルにおける混同行列のヒートマップ (n_estimators = 25)

2.7 k 最近傍法

表 8: k 最近傍法モデルにおける正解率, 適合率, 再現率, F1 値 (n_neighbors=5)

	Precision	Recall	F1-measure
クラス 0	1.00000	1.00000	1.00000
クラス 1	1.00000	1.00000	1.00000
クラス 2	1.00000	1.00000	1.00000
分類結果全体	1.00000	1.00000	1.00000
Accuracy	1.00000		

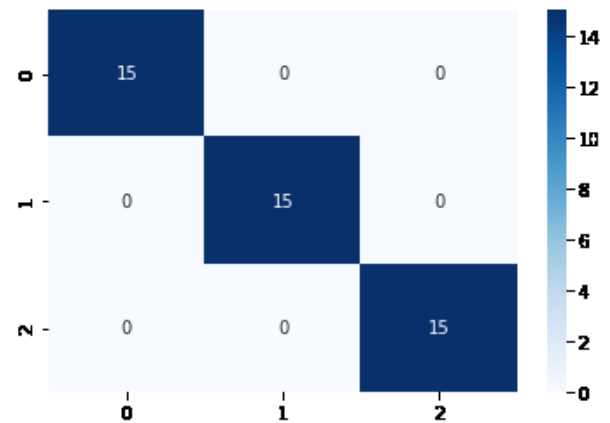


図 8: k 最近傍法モデルにおける混同行列のヒートマップ (n_neighbors=5)

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>.