

1. 実験概要

本実験では、2 目的ナップサック問題の解が与えられたときに、その解が制約を違反するかを推論する識別器の獲得を目標とする。具体的には、与えられた解について、

- 両方の目的の制約を違反: `class1`
- 目的 1 の制約を違反: `class2`
- 目的 2 の制約を違反: `class3`
- 両方の制約を違反しない: `class4`

として、各クラスを正確に識別する識別器の獲得を試みる。

本実験の実験設定は、

- ナップサック問題
 - アイテム数: 100
 - ナップサック容量: (全アイテムの大きさの総和) $\times 0.5$
- 学習データ
 - 決定変数: 各変数について 0.5 の確率で 0 か 1 に設定
 - データ数: 10,000
- 識別器
 - *k*-Nearest Neighbor (*k*-NN)
 - Random Forest (RF)

とする。なお、識別器のハイパーパラメータは、`scikit-learn` のデフォルトの設定を用いた。なお、*k*-NN のデフォルトの設定は[1]、RF のデフォルトの設定は[2]に記載されている。

学習データの教師ラベルの分布を図 1 に示す。ここで、*x* 軸のラベルは、制約を違反している目的の番号を表しており、制約を満たしている場合は “feasible” と表記している。図 1 より、このデータは、クラス分布に偏りがあるデータということがわかる。

また、より簡単な問題設定である、“目的 1 を制約違反しているか” および “目的 2 を制約違反しているか” を識別する問題についても実験を行った。両問題における学習データの教師ラベルの分布を図 2・3 にそれぞれ示す。ここでは、元の問題を “Problem 1”，目的 1 に関する問題を “Problem 2”，目的 2 に関する問題を “Problem 3” と表記する。

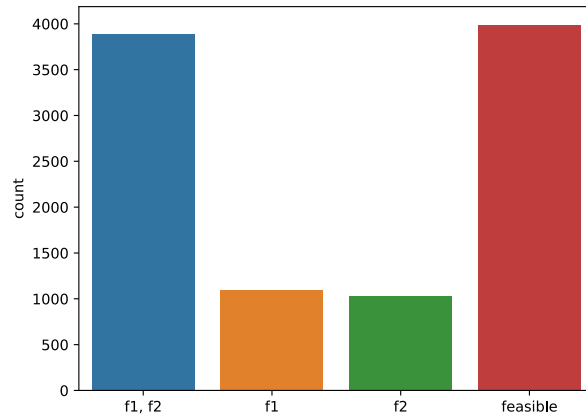


図 1: Problem 1 の教師ラベルの分布

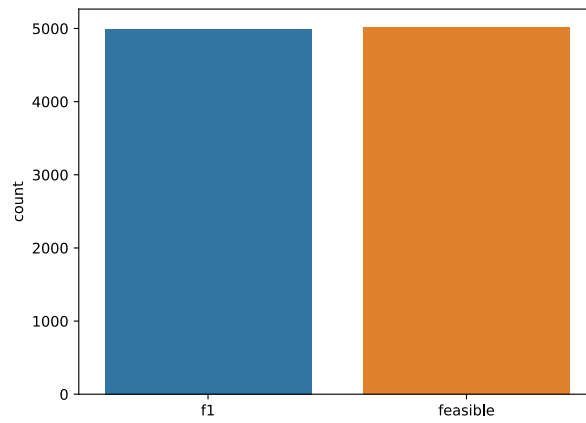


図 2: Problem 2 の教師ラベルの分布

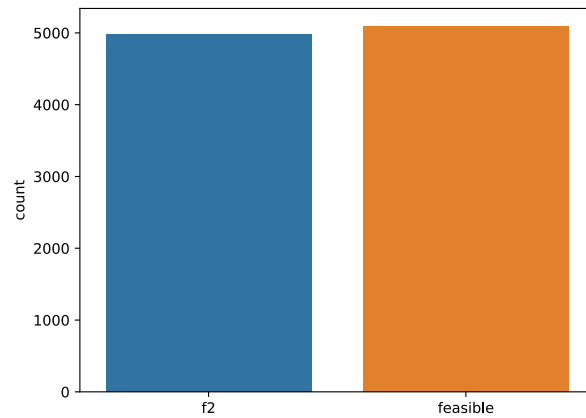


図 3: Problem 3 の教師ラベルの分布

2. 実験結果

実験結果を表 1 に示す. 表 1 より, いずれの問題においても, RF は k -NN よりも優れた識別性能を示したことがわかる. また, Problem 1 では RF であっても, 識別性能が 70%に

満たずあまりよい結果とは言えなかった．しかし，簡単化した問題では，80%を超える性能を示し，最低限実用に耐えうるレベルといえるのではないか．

表 1: 各問題で得られた識別器の識別精度(Accuracy)[%]

Classifier	Problem 1	Problem 2	Problems3
k -NN	56.34	69.98	69.34
RF	68.13	81.55	80.94

参考文献

[1]<https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html>

[2] <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>