

第 13 章

半教師あり学習

13.1 目的

半教師あり学習の手順を学びます。

13.2 Weka Collective package

Weka の半教師あり学習のパッケージである `collective-classification` をインストールすると、Explorer インタフェースに `Collective` というタブが増えます。Preprocess でデータを読み込んだあと、`Collective` タブの `Evaluation options` 領域で学習の設定を行い、`Classifier` 領域で学習アルゴリズムの選択・パラメータの調整を行います。なお、`collective-classification` パッケージで実装されている多くの学習アルゴリズムは、2 クラスの識別問題に限定されていますので、注意してください。

13.2.1 半教師あり学習における学習条件の設定

用いる学習データに応じて、学習条件の設定を `Evaluation options` 領域で行います。この領域で可能な選択は以下の通りです。

- `Cross-validation`: 交差確認法により、`Folds` で指定した分割を行った 1 つを教師ありデータとして見なしたときの、残りのデータの識別結果で評価。教師ありデータの入れ替えも可能。
- `Random split`: `Percent` で指定した分量のデータを教師ありデータとして見なし、残りのデータの識別結果で評価。
- `Unlabeled/Test set`: 教師なしデータと評価用データ（こちらは評価のため教師信号が必要）を指定して、学習と評価を行う。

実践演習 13-1

`CollectiveEM` は教科書 13.2 節で説明している自己学習の実装である。内部の識別器をいくつか入れ替え、`diabetes` データでその動作を確認せよ。

実践演習 13-2

例題 13.1 に従い、YATSI アルゴリズムを用いて iris データ、glass データなどの半教師あり学習を行え。

13.3 sklearn Label Propagation

実践演習 13-3

scikit-learn の LabelPropagation で iris データの半教師あり学習を行え。その際、教師付きデータを 5%, 10%, 20%, 30% と変化させて、性能の評価を行え。