

scikit-learn ゼミ

1.10. Multiclass and multilabel algorithms

1.10.1. Multilabel classification format

小野寺喜行

Multiclass and multilabel algorithms

sklearn.multiclassモジュール

マルチクラス分類問題とマルチラベル分類問題をバイナリ分類問題に分解することで、meta-estimatorsを実装。

Multiclass and multilabel algorithms

- Multiclass classification

各サンプルを3つ以上のクラスのどれか1つに分類

- Multilabel classification

各サンプルに複数のラベルを割り当てる

- Multioutput-multiclass classification
and multi-task classification

単一の推定器がいくつかのjoint multiclass classification
を処理する。

Multi-label classificationとmulti-class classificationを一般
化したもの。

Multiclass and multilabel algorithms

以下に示すものはmeta-estimatorsを必要としない

- Inherently multiclass:

Naive Bayes, sklearn.Ida.LDA, Decision Trees,
Random Forests, Nearest Neighbors

- One-Vs-One:

sklearn.svm.SVC

- One-Vs-All:

sklearn.svm.SVC以外の全ての線形モデル

Some estimatorsは以下のmultioutput-classificationをサポート

Decision Trees, Random Forests, Nearest Neighbors

Multilabel classification format

Multi-label classificationでのデータセット

- ・バイナリ値をもつ (n_samples、n_classes) の2次元配列
- ・各行が各サンプルを表す

例)

```
np.array([[1, 0, 0], [0, 1, 1], [0, 0, 0]])
```

↑ ↑ ↑
ラベル0 ラベル1と2 ラベルなし

Multilabel classification format

MultiLabelBinarizer

ラベルのリストのセットになっている配列をmulti-label classification用の配列に変換

```
>>> from sklearn.datasets import make_multilabel_classification
>>> from sklearn.preprocessing import MultiLabelBinarizer
>>> X, Y = make_multilabel_classification(n_samples=5, random_state=0,
...                                     return_indicator=False)
...
>>> Y
([0, 1, 2], [4, 1, 0, 2], [4, 0, 1], [1, 0], [3, 2]) ← 変換前
>>> MultiLabelBinarizer().fit_transform(Y)
array([[1, 1, 1, 0, 0],
       [1, 1, 1, 0, 1],
       [1, 1, 0, 0, 1],
       [1, 1, 0, 0, 0],
       [0, 0, 1, 1, 0]]) ← 変換後
```