Scikit-learn Seminar

1.11.4 L1-based feature selection

菊池裕紀

Selecting non-zero coefficients

- スパース性の解決策
  - ▷L1-normでペナルティをかけた線形モデルを利用
  - ▷Scikit-learnには線形モデルにデータの次元数を減らしてスパース性を緩和させる”transform”が実装されている

```
>>> from sklearn.svm import LinearSVC
>>> from sklearn.datasets import load_iris
>>> iris = load_iris()
>>> X,y = iris.data, iris.target
>>> X.shape
(150, 4)
>>> X_new = LinearSVC(C=0.01, penalty="l1", dual=False).fit_transform(X,y)
X_new.shape
(150, 3)
```

C：スパースの度合いをコントロールするパラメータ
小さいほど少ない特徴を選択する

Randomized sparse models

- L1ベースのスパースモデルの問題
  - ▷相関関係の高い特徴しか選択されない
  - ▷解決策：ランダム選択

    計画行列またはサブサンプリングデータを加えてスパースモデルを何度も再構築し、回帰が選択された回数を数える