

## 6.5 前処理ステップとモデルパラメータに対する グリッドサーチ

15t4034s 荘司 響之介

パイプラインを使うと、機械学習ワークフローのすべてのステップを1つのscikit-learn Estimatorにカプセル化することができる。  
⇒教師あり学習タスクの結果を使って、前処理パラメータの調整を行うことができる。

boston\_housingデータセットに対して、リッジ回帰を適用する前に多項式特徴量を用いる例を見ていく。  
これをパイプラインでモデル化する。

```
pipe = make_pipeline(  
    StandardScaler(),  
    PolynomialFeatures(),  
    Ridge())
```

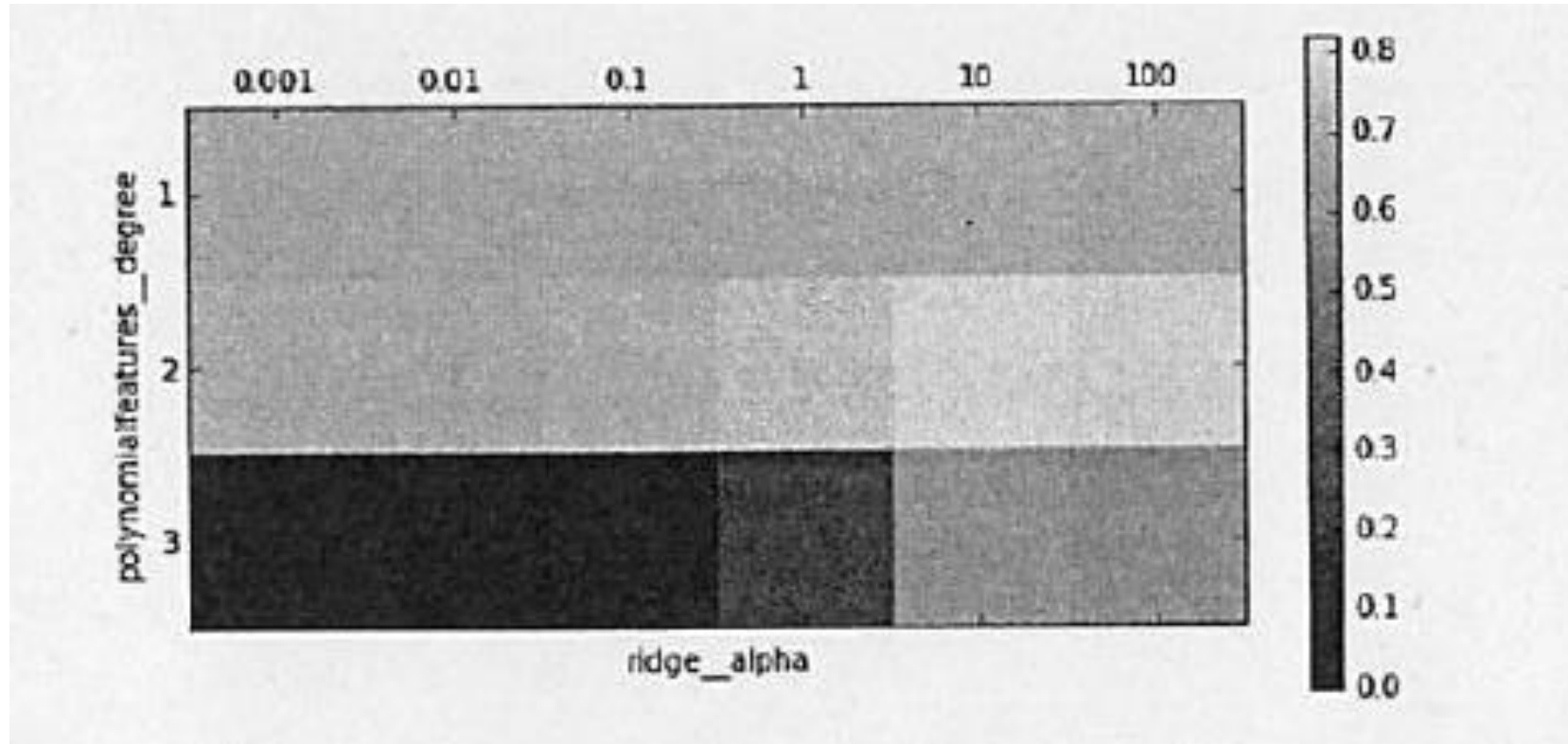
多項式の次数を選択する際、クラス分類の結果を見てdegreeパラメータを選択できれば理想的である。

パイプラインを用いることで、Ridgeのalphaパラメータと同時にdegreeパラメータをサーチすることができる。

これを行うには以下のようにparam\_gridに、双方を指定すればよい

```
param_grid = {'polynomialfeatures__degree': [1, 2, 3],  
              'ridge__alpha': [0.001, 0.01, 0.1, 1, 10, 100]}
```

グリッドサーチを再度実行する。交差検証の結果をヒートマップで可視化すると以下のようなになる。



前ページの図は、最良の結果として得られた結果にも反映されている。

polynomialfeatures\_degree:2

Ridge\_alpha:10

というパラメータを用いるとスコアは0.77となった。

(多項式を使わなかった場合、0.63)

前処理のパラメータをモデルのパラメータと同時にサーチする方法は非常に強力