

## 第 12 章

# パラメータサーチ

### 12.1 目的

学習の際のパラメータを、sklearn の Grid search および Randomized search を使って求めます。

### 12.2 Grid search

Grid search は、パラメータの可能な値をリストアップし、そのすべての組み合わせについて性能を評価して最も性能の高い組み合わせを求めます。

パラメータを組み合わせる空間をグリッドとよび、グリッドは Python のディクショナリ（辞書型オブジェクト）で表現します。試みたいパラメータの組み合わせは、ディクショナリの配列で表します。各グリッドでは、軸をディクショナリのキーに、可能な値をリスト形式の要素で表現します。

例えば、サポートベクトルマシンの「スラック変数の重み C」と「多項式カーネルの次数 degree」でグリッドを構成する場合、C の値 1, 10, 100, 1000、degree の値 1,2,3 と変化させたいときは以下のようなディクショナリの配列を作ります。

```
param_grid = [  
    {'C': [1, 10, 100, 1000], 'degree': [1, 2, 3]}  
]
```

#### 実践演習 12-1

GridSearchCV で iris データに対するサポートベクトルマシン (SVC) 識別の最適なパラメータを求めよ。なお、多項式カーネル以外に RBF カーネルも探索の対象とすること。

#### 実践演習 12-2

Grid searchCV で digits データに対する多層ニューラルネットワーク (MLPClassifier) の最適なパラメータを求めよ。

## 12.3 Randomized search

RandomizedsearchCV は、パラメータの可能な値の範囲を指定しておき、そこから乱数で値をリストアップし、そのすべての組み合わせについて性能を評価して最も性能の高い組み合わせを求めます。生成する組み合わせの数は `n_iter`（デフォルトは 10）で指定します。

パラメータが連続値の場合は、値に確率密度関数を与えて、RandomizedsearchCV にパラメータ値を生成させます。

```
param_grid = [
    {'C': scipy.stats.expon(scale=100), 'gamma': scipy.stats.expon(scale=1)}
]
```

`scipy.stats.expon` は  $\exp(-x)/scale$  に基づいて乱数を発生させるものです。

パラメータが整数の場合は、`range` 関数で範囲を与えておきます。

```
param_grid = [
    {'C': range(1,1000), 'degree': range(1,5)}
]
```

## 実践演習 12-3

RandomizedSearchCV で iris データに対する rbf カーネル SVC の最適なパラメータ、および boston データに対する RandomForestRegressor 回帰の最適なパラメータを求めよ。