

## ○ 機械学習の設計とシステムの運用

### ④ 機械学習アルゴリズムの選択

適用する問題に合わせて、教師あり学習（回帰・分類）や教師なし学習、強化学習などの各種アルゴリズムから適切なものを選びます。アルゴリズムにはそれぞれ特徴があるため、よさそうなものをいくつかピックアップして試してみます。なお代表的なアルゴリズムは第4章で紹介しています。

### ⑤ データの整形と特徴量エンジニアリング

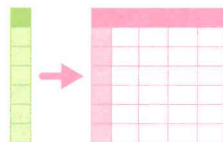
機械学習の性能向上には、どのようなアルゴリズムを利用するかも重要ですが、どのようなデータを入力するかも同じくらい重要です。アルゴリズムが受け取ることのできるデータの形式は決まっているため、別の形式で取得される場合には変換する必要があります。また機械学習では、データのの一つ一つの項目のことを特徴量と呼びますが、取得可能な特徴量をそのまますべて利用しようとするとかえって予測性能が低下する場合があります。そのため、余分な特徴量の削除や別の形式への変換、複数の特徴量を組み合わせた新たな特徴量の生成などを行い、より機械学習アルゴリズムが性能を発揮しやすくなるように調整（特徴量エンジニアリング）を行います。代表的な調整方法についてはSection14で解説します。

#### ■ 特徴量エンジニアリング

##### 余分な特徴量の削除



##### 別の形式への変換



##### 新たな特徴量の生成



### ⑥ モデルの学習

収集した学習データを利用して、機械学習モデルを学習させます。なおシステムの構築時に限らず、運用開始後も新たに収集されたデータでモデルの学習を継続することがあります。基本的な学習の方法はSection15、16で解説します。

### ⑦ 予測結果の検証と評価

予測結果が出たら検証および評価を行います。実際に利用した場合にどのくらいの性能が期待できるか知ることは、システムの運用においても大変重要です。また、必要であればさらなる性能向上のため⑤の特徴量エンジニアリングまで立ち戻って試行錯誤します。ただし、アルゴリズムは改良していくにつれ、試行錯誤の労力に対する性能向上の割合が小さくなってきます。そのため、「実用上95%の精度があれば困らない」といった打ち切りラインを定めておく必要があります。検証と評価の方法はSection17、18で解説します。

### ⑧ ハイパーパラメータのチューニング

「⑦予測結果の検証と評価」を受けて、性能向上のためアルゴリズムに指定する値の一種であるハイパーパラメータを調整します。ハイパーパラメータについてはSection19で解説します。

### ⑨ システムの運用

機械学習モデルに十分な性能が出たらシステムに組み込み、運用を開始します。ただし機械学習システムでは、運用開始後も継続的な性能の検証が不可欠です。収集されるデータの性質が変われば再度モデルの学習が必要になることもあります。また運用開始後もモデルの学習を継続している場合には、Section22のフィードバックループなどでモデルの性能が低下することもあります。

#### まとめ

■ 機械学習のシステム開発には試行錯誤が不可欠