

12

機械学習の
基本ワークフロー

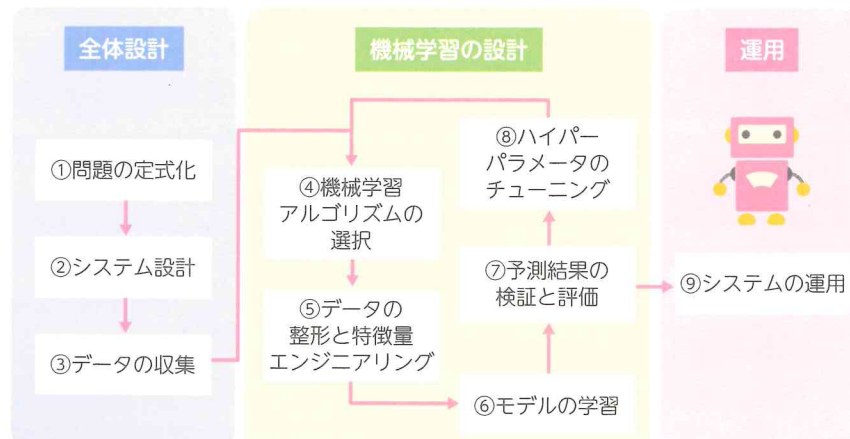
まずは、基本となるシステム開発のワークフローを見ていきましょう。スケジュールの把握や課題の見極めなど必要なことは多くありますが、トピックごとに1つずつ解説していきます。

○ 基本ワークフローと注意点

機械学習システムの開発は、通常のシステム開発に比べ、アルゴリズムの選定や機械学習の性能向上のために試行錯誤が多く、プロセス間をまたいだ手戻り（前段階に戻ってやり直すこと）が発生しやすいといえます。そのため各プロセスに費やす時間を適切に管理することが重要になります。

ただしここで重要なことは、「解決したい問題がそもそも機械学習に向いているのかどうか」を事前に見極めておくことです。機械学習により得られる予測は必ずしも正しいわけではありません。問題によってはSection04で取り上げたエキスパートシステムを利用したほうが効率がよい場合などもあるため、まずは「他のアプローチはないか」の検討を必ず行いましょう。

■ 基本ワークフロー



○ 全体設計

①問題の定式化

機械学習のシステム開発をするということは、「ネット通販で収益を増やしたい」「顧客満足度を向上したい」など何かしらの目的があるはずで、機械学習を利用して目的を達成するためには、「機械学習でどんな情報を得たいのか」を具体的に掘り下げて考えなくてはなりません。機械学習では特に、求めているものが何かによって、入出力するデータからアルゴリズムの選定に至るまでのすべてが変わるため、最初の定式化が重要であるといえます。

■ 問題の定式化

ネット通販の収益を増やすには……

→ 商品をおすすめしたい

→ 「顧客がある商品を買う可能性」を予測 ここまで具体化

②システム設計

システム設計においては、機械学習の詳細を除いた全体としてのフローを考えます。特にデータをどこから取得して、最終的にこういった形で利用するのかを先にしっかりと決めておくことで、のちのプロセスで機械学習部分での試行錯誤が行いやすくなります。

③データの収集

機械学習システムにおいて、学習や予測に利用するデータの収集は不可欠な機能です。自分が今持っているデータ以外にも、官公庁や企業が公開しているデータや、インターネットから収集したデータを活用することができます。インターネットからデータを収集する方法についてはSection13で解説します。