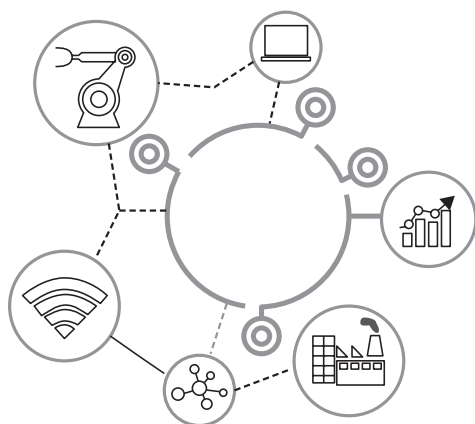


Python で学ぶ機械学習入門講座

No.1

# 基礎編

監修 / 株式会社 ROX  
執筆 / 福井 高志



コガク

---

# 目次

---

学習を始めるにあたって.....	1
第1分冊 学習のねらい .....	2
<b>第1週 機械学習とは .....</b>	<b>3</b>
1.1 機械学習とは.....	4
1.1.1 機械学習とは? .....	4
1.1.2 機械学習の種類 .....	6
1.1.3 機械学習の盛り上がりの背景 .....	7
1.2 機械学習実務の流れ.....	10
『まとめと練習問題』.....	14
<b>第2週 環境構築 .....</b>	<b>17</b>
2.1 環境構築.....	18
2.1.1 Pythonとは? .....	18
2.1.2 Anacondaとは? .....	19
2.1.3 Anacondaのインストール.....	19
2.1.4 Jupyter Notebookの使い方 .....	26
『まとめと練習問題』.....	36
<b>第3週 Pythonの基本操作①.....</b>	<b>37</b>
3.1 変数とデータ型 .....	38
3.1.1 変数.....	38
3.1.2 データ型.....	42
3.2 リスト、タプル・集合型、辞書型 .....	48
3.2.1 リスト .....	48
3.2.2 タプルと集合型 .....	53
3.2.3 辞書型 .....	57
3.3 制御構文.....	60

3.3.1 条件分岐 .....	60
3.3.2 繰り返し .....	65
『まとめと練習問題』.....	71
<b>第4週 Pythonの基本操作② .....</b>	<b>73</b>
4.1 関数.....	74
4.1.1 関数の定義と利用.....	74
4.1.2 無名関数.....	77
4.2 モジュール .....	79
4.2.1 NumPy.....	84
4.2.2 pandas .....	89
4.2.3 matplotlib .....	96
4.3 パッケージ .....	103
4.3.1 パッケージ管理ツール.....	103
4.3.2 パッケージの利用.....	103
『まとめと練習問題』.....	105
<b>STEP UP .....</b>	<b>106</b>
<b>参考文献.....</b>	<b>107</b>
<b>練習問題の解答.....</b>	<b>108</b>
<b>索引.....</b>	<b>109</b>

---

---

# 第1週

## 機械学習とは

### 【学習のポイント】

今週は「機械学習全般」についてのトピックスを広く扱います。「機械学習とは何なのか」という概念に始まり、機械学習を含む人工知能ブームの前提として、インターネットの浸透、スマートフォンの普及、SNSの隆盛の3つを取り上げます。過去からの流れに沿って概要を理解することで、より理解を深めていただければと思います。

後半では、機械学習を活用する際の実務の流れを8つに分けて簡単に紹介します。次週以降、特に第2分冊の学習の際に全体像におけるどの部分の話なのかを意識できるよう、全体の流れをよく頭に入れておいてください。

## 1.1 機械学習とは

### 1.1.1 機械学習とは？

機械学習とは、データから規則性やパターン、傾向などを学習し、それに基づき未知のものを予測、判断する技術を指します。

機械学習の有用性が今般の AI（人工知能）ブームを支えている要因の一つだと言っても過言ではありません。ただし、機械学習とは、あくまで人工知能の一分野であり、機械学習以外にも、日本語や英語などの自然言語を機械に理解させようとする「自然言語処理」など、様々な分野があります。

本講座では深層学習（ディープラーニング）については詳しく取り上げませんが、AI の一分野として機械学習があり、機械学習の中の数ある手法の 1 つにニューラルネットワークという手法があり、その拡張版が深層学習である、というおおよその関係を理解しておいてください。ニューラルネットワークについては第 2 分冊で改めて扱います。

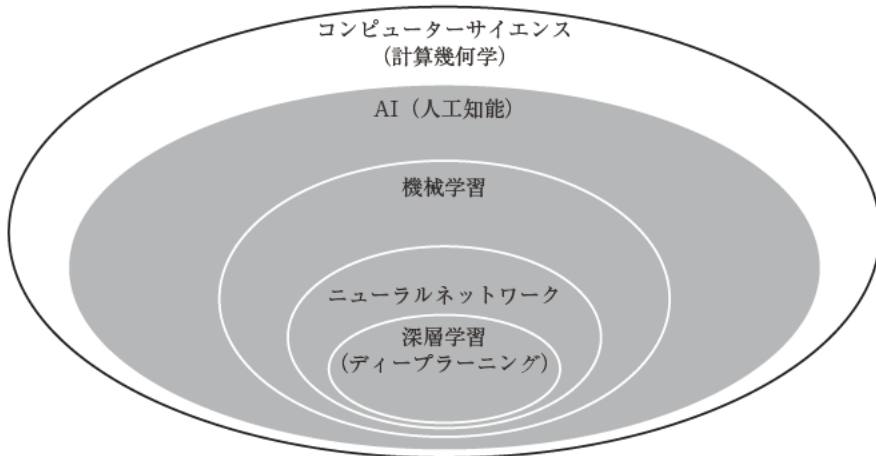


図 1.1

機械学習のステップは、大きく「学習」と「推論」の 2 つに分けられます。

1 ステップ目の「学習」フェーズでは、大量のデータを機械（コンピュータ）に投入し、そのデータに潜むパターンや傾向、ルールを機械に導き出してもらいます。

なお、データからのパターン抽出は機械が行いますが、対象のデータ、並びにパターン

を抽出するためのアルゴリズムは人間が機械に与える必要があります。

非常にシンプルな例を挙げると、過去の毎日の気温とアイスの売上を大量に機械に読み込ませれば、機械学習により「気温とアイスの関係」を導き出すことができます。それは恐らく、「気温が高いほどアイスの売上が高い（暑ければ暑いほどアイスはよく売れる）」といった内容になるでしょう。

これぐらい簡単な例であれば、人間でもすぐにパターンを読み解けそうですが、データが多量多量になればなるほど機械学習が威力を発揮します。

データに潜むパターンや傾向を導き出すメリットは、そのパターンや傾向を使えば、未知のデータについても予測や判断など何らかの対応ができることです。

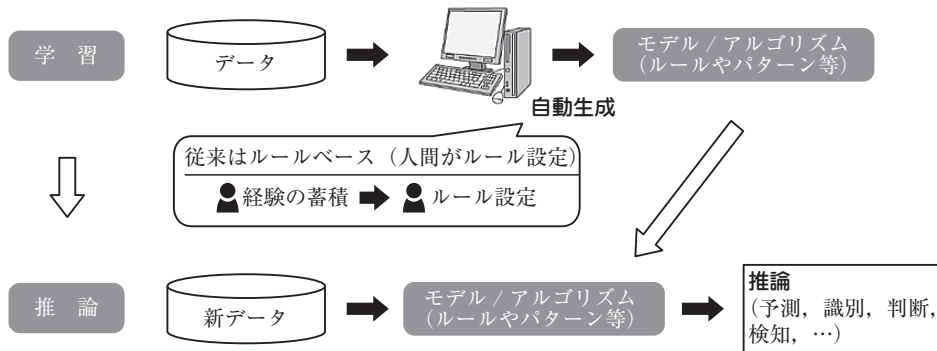


図 1.2

次に2ステップ目の「推論」フェーズです。再びアイスの例に戻ると、明日のアイスの売上というのは将来の「未知」ですが、天気予報を見て明日の予想気温が38度だと分かれば、「明日は暑そうだからアイスがよく売れそうだ」という予測ができたり「過去のパターンからXXX本ぐらい売れそうだから、多めに仕入れておこう」という判断ができたりするのです。

このように、新しいデータに対しても、学習において抽出したパターンや傾向を当てはめることで、将来の事象の予想や、判断などを行える、というのが機械学習を活用するメリットになります。

ここで注意していただきたいのは、あくまで機械学習により予測できる将来は、過去からの延長上にある将来であるということです。学習に用いたデータ（「学習データ」または「トレーニングデータ」とよびます）から導き出した傾向やパターンをベースに何かを予測するので、全く新しい未来は予測できませんし、学習データに含まれないような突発的な事象には対応できません。機械学習を含むAIは万能ではないことに注意してください。

そのような限界はあるものの、機械学習のメリットは強力で、多くの場面で使われます。特に、すでに先に触れたようにデータ量が膨大になればなるほど、機械学習の出番と言えるでしょう。人間は3次元までしか目で見たりイメージして考えたりすることができないため、それを超える多次元データとなると目視でパターンを読み解けません。しかし、コンピュータは何次元でも関係なく、大量のデータ項目（変数）の関係性を高速に計算することができます。ただし、次元が高くなるほど計算に時間はかかります。

また、機械は24時間休みなく働くことができる点でも機械学習の方が人間の手作業に比べて大量のデータを扱うのに長けています。

さらには、機械であれば長時間の作業によって集中力が落ちたり、体調不良により精度が悪くなるということもありません。ただし、機械であっても100%の精度が出ることはありませんし、常に人間よりいい精度が出せるとは限らないことには注意してください。精度の考え方、算出方法については第2分冊で扱います。

### 1.1.2 機械学習の種類

機械学習は、教師あり学習、教師なし学習、強化学習の3つに大別されます。

種 類	教師データ	概 要
教師あり学習	与えられる	アウトプットデータを予測/分類する
教師なし学習	与えられない	値やデータの類似性に基づき、インプットされたデータをグルーピングする
強化学習	与えられない（代わりに報酬やペナルティが与えられる）	与えられた環境下で、得られる報酬を最大化する行動を試行錯誤しながら学習する

#### (1) 教師あり学習

教師あり学習と教師なし学習の違いは、学習フェーズにおける教師データの有無です。教師データとは、現実のデータや人間による判別から得られた正解データのことです。

犬の写真と猫の写真を分類する機械学習であれば、それぞれの写真が犬か猫かという正解または人間による判断が教師データとなります。学習データとして大量の写真を機械に読み込ませる時に、それぞれが犬と猫のどちらの写真なのかを、教師データとして1枚1枚の写真とセットで機械に教えるのです。気温からアイスの売上を予測する機械学習であれば、各気温における売上の実績データが教師データとなります。

教師あり学習は、主に「回帰」と「分類」に利用されます。

回帰は連続値を予測するもので、株価の予測や商品販売数の予測などに用いられます。分類はカテゴリーを判別するもので、代表的な利用例として画像の分類、例えば手書き文字が 0～9 のどれに当たるかの判定や、人物の写真から性別や年齢層を判定する等が挙げられます。

## (2) 教師なし学習

Amazon 等の EC サイトで商品を購入した際に別の商品を勧められた経験をお持ちの方も多いと思いますが、ここには教師なし学習の手法が使われています。商品の閲覧履歴や購買履歴といったデータを元に顧客をグルーピングし、ある顧客グループ内の多くの人が A という商品をよく買っているにも関わらず、その顧客グループの中に A をまだ買っていない人がいれば、「あなたには A がオススメです」などとレコメンドを出すことで購買意欲を刺激するのです。

顧客のグルーピングにおいては、どう分けるのかに正解はありません。つまり教師データは存在しないということです。教師データのある学習データを読み込むのではなく、機械が各データの類似度から似ている度合いを判断し、近い人をグルーピングします。何を以って類似度を測るかという基準は、人間が設定します。

## (3) 強化学習

世界最強の棋士を破って話題となった囲碁 AI には、強化学習の手法が取り入れられています。

強化学習では、教師なし学習同様に教師データが与えられませんが、ある環境における「行動の選択肢」と連続した行動の結果に対する「報酬」が与えられます。その環境において、一連の行動を通じて得られる報酬を最大化するように学習を進めるのが、強化学習です。

本講座では、教師あり学習と教師なし学習を扱います。教師あり学習／教師なし学習それぞれに複数の手法（アルゴリズム）がありますが、その中でも主な手法について第 2 分冊で学習していきます。

### 1.1.3 機械学習の盛り上がりの背景

ここで、機械学習の利活用が進んだ背景として、ここ 20～30 年に起こった世の中の変