

第0章

はじめに

のつけから失礼しますが、機械学習をなんのために勉強しますか？ おもしろそうだから、という人もいるでしょうが、ほとんどの人は現実の解きたい問題に应用するためですね。

何かを応用できる形で身につけたいとき、ひたすら「こういうときはこうする」を覚えるという手があります。実際、教科書や問題集に載っているような問題なら、「こういうときはこうする」でだいたい解けるでしょう。

しかし現実の問題はパターンもデータも千差万別、算数のドリルのようには解けません。しかも本書でのちほど説明するとおり、機械学習には正解なんかありません。100点満点の答えが存在しないからこそ、「なぜその方法で解いたのか」「なぜそんな計算ができるのか」は重要です。

もう少し具体的に言うと、機械学習そのものやそのモデル・アルゴリズムがその形になっている成り立ちや動機を知り、「機械学習がそんなことをしたい、してもいい理由」を把握することで、今解きたい問題に機械学習を使うにはどうしたらよいか、どこを変えると性能を上げられる可能性があるか、そもそも機械学習を使わないほうがいいのか(!)という判断ができるようになります。

とはいえ、機械学習のモデルは無数にあり、すべてのモデルがそれぞれの理由を持っています。そのすべてを紹介することは不可能ですが、機械学習の理由を組み立てるパターンは実はそんなに多くありません。この「機械学習の理由を組み立てるパターン」が本書のテーマ、機械学習の「理屈」と呼んでいます。

機械学習の「理論」とどう違うの？ という疑問もあるかもしれませんね。機械学習の理論とは、体系付けられた知識と、それをもとに「こうあるべき」を分析したものです。平たく言うと、「機械学習は何ができるのか、何ができないのか」を明らかにするのが理論の仕事であり、機械学習を安心して使うためにとても重要です。「機械学習でそうすべき理由」は理論で説明できますが、本書のターゲットである「そうした理由」は説明できません。

この本は機械学習を身につける早道ではありませんが、機械学習をこれから勉強しようとしている人、機械学習を勉強してみたけど「なぜこんなことをするんだろう」というモヤモヤを抱えている人には、機械学習の理由や理屈という「急がば回れ」はきっとよく効くと思います。

0.1 本書の対象読者と構成

本書は「数式がなくてもわかる本」ではありません。機械学習の理解と応用には数学が重要であり、本書でも「機械学習の理屈」に必要な遠慮なく数式を使います。数式を飛ばして読んでも得るものがあるように書いているつもりではありますが、数式もしっかり読んでもらうのが一番です。

したがって本書の主要な想定読者は、本書で用いる数学である線形代数（ベクトル・行列）と解析（微積分）についてひととおり学んだことのある人となります。本書で使う線形代数と解析の知識と記法は、巻末の付録に簡単にまとめています。

また機械学習には（モデルにもよりますが）確率の知識も必要です。こちらは確率の基本的な考え方（第2章）、連続確率（第3章）、そしてベイズ確率（第5章）の3章に分けて手厚く解説しています。

線形代数たちとの扱いの違いは、確率が「一見簡単に見えるが、高校までで学んだ『確率』とは本質的に異なる」という事情を持つためです。統計や機械学習において重要な役割を果たすベイズ確率や連続確率は、「高校までの数学に出てくる確率」の延長で考えようとする気持ち悪くてモヤモヤするでしょう。確率を「起きるかもしれないことがら」を数値で表現する統一的な枠組みという近代的なモデルで理解することで、初めて確率の理屈を納得できます。そのため確率については特に知識を仮定せず、かつ本書の方針に従って「確率とは何か、なぜそのようなものを考えたいか」から解説しています。

機械学習や統計で用いられる分布やモデルのカタログ的な本も多くありますが、本書はそのような網羅性はなく、紹介する機械学習のモデルは主に初歩的なものに限られます。流行りの深層学習についても、機械学習の枠組みの中でどのような位置づけにあるかなどは紹介していますが、具体的な手法については触れません。

本書の各章の内容を紹介します。

第1章では、機械学習ともっとも重要な考え方であるモデルとは何者で、何ができて何ができないかを説明します。また人工知能と深層学習についても、機械学習との関係を中心に簡単に解説しています。第2章（確率）と第3章（連続確率と正規分布）、そして第5章（ベイズ確率）は上述のとおり確率その考え方から説明しています。第4章は機械学習のもっとも基本的なモデルの

ひとつである線形回帰を解説します。第6章ではその線形回帰をベイズ化することで、ベイズ確率の大きなメリットのひとつを実感してもらいます。第7章は、機械学習の多くの応用で使われる分類問題について、その代表的なモデルのいくつかとともに紹介します。第8章では、最適化法と言われる関数の最小値を探索する手法全般を簡単に紹介し、ロジスティック回帰（7.4節）を実際にその手法のひとつで解きます。機械学習を含めた多くの科学分野では、問題を解の良さを表す関数に書き換えて、最適化の問題に落とし込む枠組みで解かれることがとても多いです。第9章は機械学習のモデルを選択する方法です。この章までに見るとおり、機械学習の解は選んだモデルによって決まるため、良いモデルを選ぶことは機械学習でもっとも重要なことのひとつです。最後に第10章で本書のまとめと補足を行います。

0.2 謝辞

本書の執筆にあたっては、伊藤徹郎さん、平田智章さん、山田高大さん、そして江原遙さん（静岡理科大学講師）を始めとした多くの方にご協力いただきました。いただいたレビューのおかげで間違いやあいまいさを減らし、わかりやすさを改善することができました。本当にありがとうございました。

機械学習ことはじめ

機械学習や人工知能というと「データから勝手に学習して、使えば使うほど賢くなる」というイメージがあるかもしれません。実際、そう謳^{うた}っている記事やサービスを目にしたこともあるでしょう。

しかし、夢を打ち砕いてしまったらたいへん申し訳ないですが、機械学習は『学習』しませんし、現実の人工知能は勝手に賢くなったりしません。

機械学習が「データから学習する」というときの「学習」は機械学習の専門用語であり、一般動詞の『学習する』、つまり新しい知識を身につけたり、今までできなかったことができるようになることではないのです。

本章は機械学習の正しい理解と、多くの科学分野で重要な概念である「モデル」について解説します。また、最近ブームの深層学習と人工知能について、機械学習とどのような関係があるか簡単に紹介します。

1.1 機械学習とは

機械学習の「機械」は、印刷機やポンプや工場に並んでいるようなガッツリした機械などではなく、コンピュータを指しています。

では、機械学習の「学習」のほうはなんでしょう。

日常的な意味での『学習』とは、本を読んだり問題を解いたり誰かに教えてもらったりして、知らなかった知識や技術を理解し覚え身につけ、今までわからなかったことがわかるようになり、できなかったことができるようになることでしょう。この本を読むことも立派な『学習』のひとつです。

しかし機械学習とは、コンピュータが何かを理解して、できなかったことができるようになる技術ではありません。具体的な機械学習の手法の紹介はのちの章に譲って、この節では「機械学習とは本質的に何か」を身近なテレビの例を使って説明しましょう。

ほとんどすべてのテレビやディスプレイは、赤・緑・青の色の強さを調整できるようになっています（図1.1）。

例えば青のつまみを数値が増える方に動かすと、青の成分が強調された表示になります。これは「テレビに映る色を、実際に目にしたときに感じる色に近づけたい」という問題を解決するための仕組みです。

例えばテレビに映っている青空が少しくすんで見えるなら、画面を見ながら各色の強さを調整することで自分の感じる色に近づけられます。見たこともな

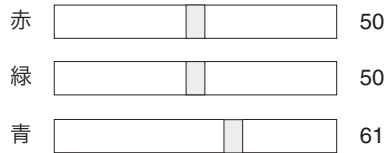


図1.1: テレビの色調整機能

いものを映しても本当の色はわかりませんから、調整には実際の色を見知っている映像が必要です。シチュエーションや質感によって色の見え方も異なるでしょうから、いろいろな種類の映像をいくつか用意しておいたほうがよさそうです。確実に期すなら、調整用とは別に確認用の映像も用意したいところです。その映像の色の確認を別の人にしてもらえればさらに安心でしょう。

実はこれ、まさに機械学習の枠組みそのものなのです。

機械学習の枠組みでは、解きたい問題を数値で扱えるようにしたものを**数理モデル**あるいは単に**モデル**と言い、モデルが持つ調整用の数値を**パラメータ**、パラメータの調整に使うデータを**訓練データ**（**トレーニングデータ**）、モデルが正しく働か確認に使うデータを**テストデータ**と言います。そして「データを使ってモデルのパラメータを適切に調整する」ことを**学習**あるいは**訓練**と呼びます*1。

先ほどの例では、「テレビの色調整の仕組み」がモデル、「赤緑青の各色の強さを指定する数値」がパラメータ、「実際の色を見知っている映像」のうち調整に使う映像が訓練データ、確認に使うのがテストデータ、そして「自分の感じる色に近づくように各色の強さを調整」が学習にあてはまります。

ただ、思ったとおりの色に近づいたかどうかを人間が判断しているため、枠組みこそ同じですがまだ「機械」学習ではありません。あとは近づき度合いを数値で表し（誤差や目的関数）、調整を手順化することで（アルゴリズム）、立派な機械学習になります。

こんな色調整の仕組みが機械学習の枠組みとは思えない、と反論したくなる

*1 このようなパラメータを持つモデルを特に区別する場合は**パラメトリックモデル**と言います。一方、本書では扱いませんが、「ノンパラメトリックモデル」もあります。しかしこれはパラメータがまったくないという意味では必ずしもありません。パラメトリックモデルの表現可能な範囲はパラメータの動く範囲に限られ、データには依存しません。それに対してモデルの表現可能な範囲がデータに依存するとき、ノンパラメトリックと言います。

かもしれません。たしかにこれでは「できなかったことができるようになる」というより、「もともとできる範囲内で、よりうまくいくようにする」であり、『学習』っぽくないです。

機械学習の始まりはたしかに人間の『学習』をコンピュータで実現する研究でした。しかし『学習』の完全な実現はあまりにも難しすぎたため、『学習』そのものではなく、『学習した結果としてできるようになること』を再現する方向へ発展します。特に「モデルを決めて、データに合うパラメータを探す」という汎用性の高い枠組みが成功を取って、現在の機械学習の主流となりました。こうして機械学習は、一般的な意味での『学習』からは遠くなってしまいました。

機械学習では、まず実際に解きたい問題に合わせて、モデルをうまく選んだり、場合によっては自分で作ります。解ける問題やできるようになることの範囲は、最初に選んだモデルで決まってしまうため、モデル選びはとても重要です。また、どれほど素晴らしいモデルだったとしても、パラメータ調整の効率が悪く、現実的な時間で解が求まらなければ意味ありません。優れたアルゴリズムや十分なマシンパワーも必要です。

そして、もうひとつ重要なものがデータです。機械学習は見たことないものが苦手ですから、賢い機械学習のためには一般に網羅的で質の良いデータが大量に必要です。そんなデータを集め整備するのは主に人間の仕事です。したがって機械学習（後述しますが人工知能も含みます）が勝手に賢くなるのはとても難しいです。

とはいえ機械学習は、数値の計算しかできないコンピュータに、人間のするような高度な仕事を任せられる（可能性がある）、素晴らしい技術です。機械学習にできないことを期待せず、機械学習にできるすごいことにきちんと目を向けていきたいですね。

1.2 モデルとは

先ほど、機械学習ではモデルが重要であるという話をしました。しかし、モデルとはいったいなんでしょう。「解きたい問題を数値で扱えるようにしたもの」では、わかるようでわかりません。

モデルとは、直訳の模型という意味のとおり、「何かの偽物」であり、本物そのものではありません。しかし偽物ならなんでもよいわけではありません。

例えば人体模型は人間の体の偽物です。本物の人間の体と見間違ふ心配がまったくないくらい似ていません。人体模型はパーツを取り外して手にとって見るなど、人体の内部構造がどうなっているか観察できるという点で役に立ちます。本物の人間の体でそんなことは軽々しくできません。

このように本物ではないが、特定の目的において（ときには本物より）役に立つ偽物がモデルと呼ばれます。ほとんどのモデルは何かひとつふたつの目的に特化しており、それ以外の部分を似せることは最初からあきらめています。

本物が持ついろいろな側面のうち、どれをどのように似せるかによってさまざまなモデルが作られます。つまり、モデルを設計したときに、そのモデルがなんの役に立つかはおよそ決まります*2。そのため、モデルが役に立たないと感じる場合は、モデルが悪いのではなく、間違ったモデルを選んでしまったためであることがほとんどです。人体模型に「こいつ、歩かない！」と文句言ってもしかたないですね。歩いてほしかったら、例えば2足歩行ロボット（これも人体のモデルのひとつ）を選ぶべきだったのです。

ちなみに、機械学習の「モデルを決めて、データに合うパラメータを探す」という枠組みも、人間の『学習』のうち、結果を見て行動を修正するという一面を真似したモデルです。この枠組みはとても役に立ちますが、人間の『学習』の「できないことをできるようにする」という機能を期待するのは、人体模型に歩くことを期待するくらい間違っています。

モデルのこのような性質は、複数あるモデル候補の中からどれを選ぶかが重要と教えてくれます。このことは第9章のモデル選択で具体的に解説します。

もうひとつ、ここまで何度か使った**アルゴリズム**という言葉についても触れておきましょう。一般向けの解説などでも「機械学習のアルゴリズムを使って～」といった形でよく見かける言葉ですよね。

モデルは問題を定式化（数値で扱えるようにすること）したものでした。それに対してアルゴリズムとは、問題を解くための手順のことです。後に紹介する線形回帰モデルでは行列計算で、ロジスティック回帰モデルでは勾配法と呼ばれる方法でそれぞれ問題を解きます。

本書では各モデルに対して1つのアルゴリズムしか紹介しませんが、数学の問題の解き方が1通りではないのと同じように、モデルに対するアルゴリズム

*2 モデル設計者の意図していなかった使い道を見つけてもらえることもあります。

も1通りではありません。同じモデルでもデータの中身や量によって適切なアルゴリズムは変わることがあります。

本書でこのあと何度か繰り返し説明するように、機械学習では問題に適したモデルを選ぶことが重要です。しかし、どれほど素晴らしいモデルでも解けなかったら意味がありません。したがって、モデルとデータに適切なアルゴリズムを使うことも同じくらい重要です^{*3}。

1.3 深層学習とは

これから機械学習の勉強を始めようという人は、もれなく**深層学習**（ディープラーニング）と人工知能にも強い興味を持っているでしょう。

さてその深層学習とはなんでしょう。「深層学習とは人間の脳細胞を再現するもの」的な紹介も見かけますが、それは残念ながら間違いです。深層学習は脳に似せることをまったく目指していません。一般的な深層学習と脳のもっとも重要な関係は、深層学習が脳についてわかっていることのひとつを参考にしたモデル（ニューラルネットワーク）から発展した、ということです^{*4}。

深層学習を一言で説明すると「ある特徴を持つ機械学習」です。深層学習は機械学習以外のものではなく、機械学習の代表的な枠組みである「モデルを決めて、データに合うパラメータを探す」に、深層学習も当てはまります。では、その深層学習の特徴とはなんでしょう。

機械学習の枠組みの例として挙げた色調整では、パラメータは赤（R）、緑（G）、青（B）の3個でした。実際のテレビやディスプレイの色や見た目の調整機能には、さらに明るさやコントラスト、色温度などのパラメータがあったりします。さらに高機能なテレビでは、動きの多いシーンとか暗いシーンとか、シーンごとの調整を指定できるものもあったりします。このように、色々なケースを考えてパラメータを増やすほど高機能になり、できることが増えていきます。

この「パラメータを増やすほど高機能になる」という性質は機械学習全般に

^{*3} とはいえ、どちらが主かといえば明らかにモデルなので、「機械学習のアルゴリズムを使って～」という表現には違和感があります……。

^{*4} 現在も神経科学（脳に関する研究）から発想を得た注意機構などを深層学習に組み込む試みは継続して行われています。

もまったく共通する話です。これを突き詰めることで「パラメータがとてつもなく多い、超々高機能な機械学習」が構築できるかもしれません。しかし、パラメータの意味を考えながら増やしていくのは限界があります。色調整のパラメータを 1,000 個考えるのは大変です。

パラメータの意味を考えていたらパラメータを増やせないなら、いっそ意味は捨ててしましましょう。まず問題を「入力 X_i たちから適切な出力 Y_j たちを生成する」というシンプルな形に整理します。色調整の場合、入力された画像データから実際にディスプレイに表示する情報を生成することになりますから、 (m, n) の位置のピクセルの色データ（画像の要素ごとに赤緑青の強さを表した数値） (r_{mn}, g_{mn}, b_{mn}) のそれぞれが入力 X_i にあたるでしょう。画面の大きさを簡単のため $2,000 \times 1,000$ ピクセルとすると、 X_i たちは $2000 \times 1000 \times 3 = 600$ 万個になります。同様に出力 Y_j にも出力した画像のピクセルの色データ (R_{mn}, G_{mn}, B_{mn}) を当てることにしましょう。

そして入力 X_i に対して出力 Y_j が次のように決まるモデルを考えます。

$$Y_j = \sum_i w_{ij} X_i \quad (1.1)$$

w_{ij} がこのモデルのパラメータで、（入力の個数） \times （出力の個数）個あります。 X_i, Y_j とともに 600 万個ずつでしたから、パラメータ w_{ij} の個数は 36 兆個になります。これで「パラメータがとてつもなく多い機械学習」ができました（「超々高機能」かどうかはわかりません）。

まだまだパラメータを増やしたい？ では Y_j の代わりに Z_k を新しい出力にして、 Y_j と Z_k の間に次の関係を入れましょう。これだけでパラメータの個数は倍になります*5。

$$Z_k = \sum_j v_{jk} Y_j \quad (1.2)$$

(1.2) のような関係をさらに重ねれば、いくらでもパラメータを増やせます。中間の Y_j の個数を適当に増やしてもよいでしょう。

こうした個々には明確な役割のない莫大なパラメータを持つモデルが深層学

*5 (1.1) を (1.2) に代入すると、 Z_k が X_i の線形結合となりパラメータが減ってしまうので、実際の深層学習では Y_j を非線形関数（活性化関数と呼ばれます）で変換します。ここではこの方法を重ねることによってパラメータをいくらでも増やせることを覚えてもらうのが目的なので、活性化関数は省略しています。

習の特徴です。この特徴は深層学習と呼ばれるモデルのほぼすべてに共通し、かつ通常の機械学習のモデルにはありません*6。

機械学習ではパラメータを増やすほど最適なパラメータを探すのは難しくなり、データも多く必要になるため、そのような莫大なパラメータを持つモデルの学習にはリソースや安定性などのさまざまな問題が生じます（4.5節の過学習など参照）。深層学習とは、そうした問題を解決するためのさまざまな技術の集合とみなすこともできます。

また、パラメータが意味を持たないということは、モデルが期待どおりに動くかどうかかわからないということでもあります。例えば(1.1)から作られるモデルは「画像を入力したら、画像を出力する」ということしか決まっています。つまり、このモデルが色調整として機能するかどうかは、用意した訓練データでモデルがうまく学習できるかで決まり、うまく動くモデルを見つけるためには膨大な試行錯誤が必要となります。

このようになかなか一筋縄ではいかない深層学習なのですが、良いモデルを運良く見つけられた場合には、従来の機械学習では太刀打ちできないレベルの機能や性能が得られます。

1.4 人工知能とは

今はたいへんな人工知能（AI、Artificial Intelligence）ブームです。人工知能に興味があつて機械学習を勉強し始めた人も少なくないでしょう。ではそんな人工知能と、機械学習や深層学習はどのように関係しているでしょう。

人工知能には大きく分けて2種類、「強い人工知能」と「弱い人工知能」があります。

強い人工知能は、人間の頭脳の働きまたは仕組みを機械で再現するもの、あるいはそれを目的とした研究のことであり、特に汎用人工知能（AGI、Artificial General Intelligence）とも呼ばれます。強い人工知能は特定の目的に偏らないさまざまな判断や思考が期待されていますが、実現はまだまだ遠いです。

マンガや映画にはドラえもんや鉄腕アトム、攻殻機動隊のタチコマ、ターミ

*6 この性質は深層学習の特徴であつて、定義ではありません。一般的な深層学習の定義は、誤差逆伝播法で勾配を計算することで深く深いニューラルネットワークになります。

ネーターのスカイネット、スター・ウォーズのドロイドといった、人間と会話し、臨機応変に問題解決し（あるいは問題を起こし）、ときには喜怒哀楽も表したりする、まるで人間のような自律的なロボットやアンドロイドが出てきます。これら物語の人工知能はすべて強い人工知能に分類できます。

強い人工知能をいきなり実現するのは難しいため、人間の知能が実現する内容を分割して、それぞれ解決していくというアプローチがあります。そのような分割された部分のひとつ、「人間の判断が関係する、具体的な個々の問題を解くプログラムや研究」が**弱い人工知能**と呼ばれます。

弱い人工知能には多くの実現例があります。例えば「将棋を指す」「ゲームのキャラクタを人間らしく動かす」「写真から文字を抜き出してテキストに起こす」「人間とそれっぽい会話をする」などそれぞれの問題に特化したプログラムはすべて弱い人工知能に分類できます。将棋 AI で写真から文字を抜き出すことはできないように、弱い人工知能はそれぞれ専門とする問題以外は解けません。弱い人工知能を見て「人間に勝てる、勝てない」という言葉が出てくることはあっても、「まるで人間のよう」という言葉が出てくることはほとんどないでしょう。

機械学習はこの弱い人工知能を実現するための手法のひとつとして誕生しました。ただし弱い人工知能が必ず機械学習を使うわけではありません。現在は機械学習（深層学習含む）をまったく使わない弱い人工知能はほとんど見られなくなりましたが、20～30年ほど前の主流は、専門家の判断をルールとして書いて判断を自動化するもので、エキスパートシステムと呼ばれていました。将棋を指す人工知能も、機械学習を使った Bonanza が世界コンピュータ将棋選手権大会で2006年に優勝するまでは、ルールベース（評価関数を手作業で作る）が主流でした。会話をする人工知能もルールベース（問い合わせに対する対応を人間が決める）が基本であり、すべてが機械学習などに置き換えられるのは先の話でしょう。

一方の強い人工知能は研究ごとにアプローチがまったく違うため、機械学習を使うものもある、くらいのことしか言えません。最近は強い人工知能にも深層学習の応用が増えています。

このように強い人工知能と弱い人工知能はそのできることも実現度合いもまったく異なる存在ですが、一般にはどちらも単に AI と呼ばれて区別されません。人工知能を巡る報道や宣伝では、この混同がうまく利用されたり、そも

そも違いが理解されていないように見えることも残念ながら多いです。例えば、「将棋を指す AI」は将棋を指すこと以外にはできない弱い人工知能のことで、ところがこれを「AIが将棋を指す」と言い換えるだけで、他のこともできそうな強い人工知能っぽさが出てきます。これを繰り返すことで、世の中に強い人工知能があふれているような錯覚を生じさせられます。

こうした錯覚に引きずられて、近い将来に人工知能が反乱を起こして人間を支配したり、人間の仕事がすべて人工知能に取られるかのような危機感を煽ってくる言説もありますが、幸か不幸かそれが可能な強い人工知能の実現はまだまったく目処は立っておらず、実際に実現や応用されている人工知能は基本的にすべて弱い人工知能です*7。

強い人工知能が実現し、機械が人間に迫いつき、社会が変化してしまう日を**シンギュラリティ**（特異点）と呼び、それが30年後に実現すると主張する人たちもいます。このシンギュラリティは、科学技術が指数関数的に進歩するという経験則にもとづくものであり、技術的な裏付けから提唱されているものではありません。しかも、その経験則の代表格だったムーアの法則（半導体のトランジスタ数は18ヶ月ごとに2倍になる）はすでに破れたとされています。思考実験としては興味深いですが、まるで決まった未来であるかのように論じるのは控えてほしいところです。

残念ながら、現在の機械学習や深層学習の技術をそのまま伸ばしていった先に強い人工知能はまだありません。物語の人工知能の実現にはまだまだ技術的な課題が山積みですが、人間の脳だって物理的なデバイスとして存在している以上、いつかは機械で再現できるだろう、と実は筆者も信じています*8。そんな未来が本当にやってくるまでは、機械学習や深層学習で弱い人工知能を作って役立てたり楽しんだりしましょう。

*7 弱い人工知能によって置き換えられてしまう人間の仕事もあるかもしれませんが。しかし弱い人工知能は特定の問題を解くことに特化したプログラムであり、誰かがそれを解くことを望まない限り存在しません。するとそれは自動車の発明と普及で馬車がなくなるような、技術の進歩によって起きる普遍的な問題であり、人工知能に固有の話ではありません。

*8 生きているうちに実現するかはわかりませんが……。