

15 モデルの作成と学習

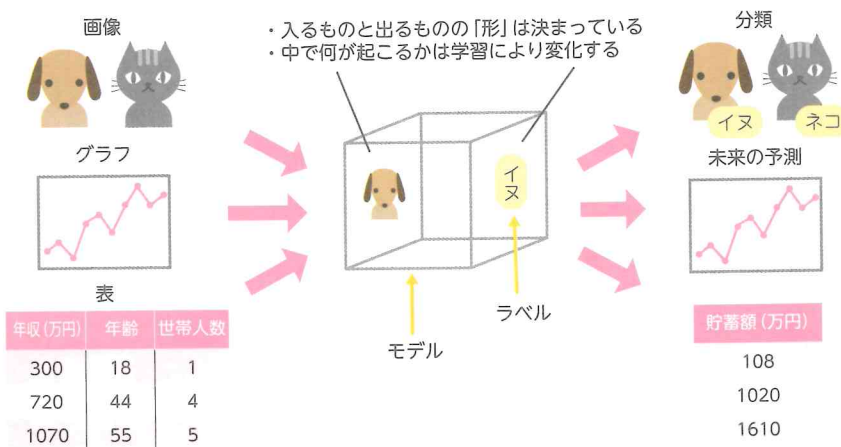
機械学習では、適用する問題に合わせてアルゴリズムを選択し、モデルを作成します。このセクションでは、モデルの作成と学習の方法について、理解を深めていきましょう。

○ モデルとは

機械学習における**モデル**とは、入力されたデータからある出力（入力データに対する分類や予測）を導き出すための数理モデルのことです。このことは、「**何かを入ると何かが出てくる箱**」をイメージすると、わかりやすくなります。箱はそれぞれ大きさや入口の形が異なり、それによってどんなものを入れられるかが決まります。同様に、箱から出てくるものの形も決まっていると考えてください。そのようなモデルを作成するためには、入出力するデータがどのようなものであるかを最初に決定しなければなりません。

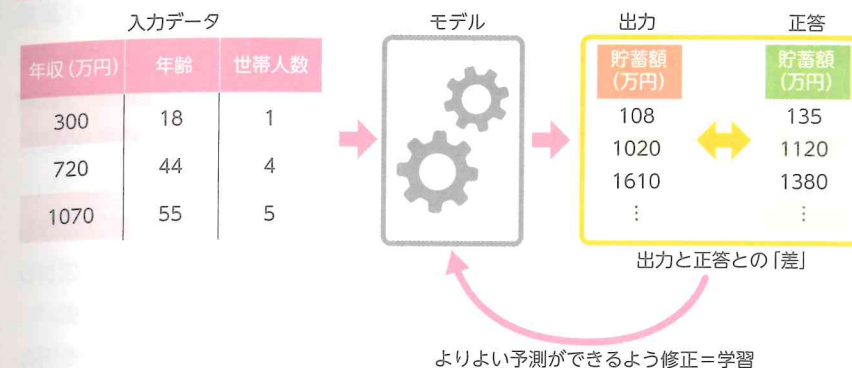
ちなみにこの「箱」は、数学における「**関数**」にあたります。機械学習において、アルゴリズムの中で行われていることは関数の計算なのです。

■ モデルは「関数」



モデルを作成した時点では、その内部での処理（関数の計算）はでたらめな場合がほとんどです。そのため何が起きるか見当は付きません。イヌの画像を入力してもネコと分類されてしまう可能性もありますし、未来のグラフを予測させても全く見当はずれな予測をするかもしれません。このような状態のモデルに対し、よりよい出力ができるよう修正することを「学習」と呼びます。

■ ランダムな処理を修正する



このように正答データとの差を小さくするように学習させる手法は教師あり学習です。なお教師なし学習は、手法により学習の過程が大きく異なるため、第4章で取り扱います。例として、ある人の年収から貯蓄額を予測する機械学習アルゴリズムを考えましょう。この場合、学習データとは「年収〇〇万円」といった調査対象のデータのことで、正答データはその人が「実際に貯蓄している金額」のことで

以上を踏まえて、年収を入力データとした機械学習モデルを考えます。入力データとなる年収を横軸、正答データとなる貯蓄額を縦軸にしたとき、このデータは2次元グラフに点として描けます（P.72下図）。すでに何となく「年収と貯蓄額は比例するだろう」と予想できますが、グラフを確認しても、やはりそのような傾向を持った直線を確認することができます。

このように、比例するデータに対しては一般的に直線モデルが有効です。直線は傾きと切片の2つのパラメータで処理が決定されるので、学習データをフィット（適合）するようなパラメータを探すことでよいモデルを作成することができます。