

06

教師なし学習のしくみ

教師なし学習は、データの構造や法則をアルゴリズムによって解析し抽出する機械学習の手法です。教師あり学習とは異なり、人間が正解を教えることなく学習が進行する点が特徴です。

教師なし学習は「データの特長をとらえる」

教師なし学習(**unsupervised learning**)は、与えられたデータの本質的な構造や法則をアルゴリズムが自動的に抽出する機械学習アルゴリズムです。教師あり学習では人間が教師となり、分類タスクであればデータのカテゴリ名、回帰タスクであれば具体的な数値といったような正答データを、学習データとセットにしてアルゴリズムに与えます。このように正答データを与えることからわかるように、教師あり学習のゴールは「未知のデータに対して正しい回答をする」ことです。対して、ここで解説していく教師なし学習では、正解データを用意せず学習データのみをアルゴリズムに与えます。そんな教師なし学習のゴールは「データの特長をとらえる」ことにあります。

教師あり学習と教師なし学習の違い

?に入るのは
「65」だ!



文系と理系の2つに
分けられそうだ!



教師あり学習

英	数	国	理	社	合格高校の 偏差値
80	60	70	65	75	60
90	80	75	70	80	?
⋮	⋮	⋮	⋮	⋮	⋮

正解データ

教師なし学習

英	数	国	理	社	
80	60	70	65	75	
90	80	75	70	80	
⋮	⋮	⋮	⋮	⋮	

正解データがない

人間はいくつかのものを見るとき、意識しないままそれぞれの特徴をもとに「区別」しようとします。たとえばあなたが、図のように並べられた野菜や果物を見たとき、たとえ名前を知らなかったとしてもただ漠然と眺めることはしないはずです。色で分けてみたり形で分けてみたりし、どんなグループを作ればその状況をうまく説明することができるのかを考えるでしょう。そして野菜や果物を色で区別することによって初めて、「ここには6色の野菜や果物があります」といった風にまとめることができます。教師なし学習ではこの人間の「特徴をとらえる」能力をアルゴリズムにより再現することを目指しているのです。

野菜や果物を区別する

ものを「区別する」

区別すると「特徴をとらえられる」

