

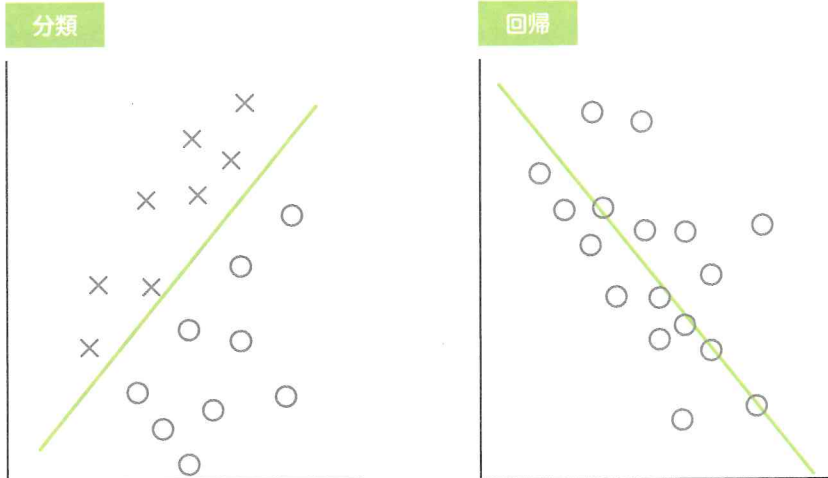
○ 機械学習が扱う問題（分類と回帰）

機械学習の問題は、大きく**分類**と**回帰**に分けられます。まず分類は、どのデータがどの種類に属するかを見ることが目的です。入力されるデータがいくつかのグループに分けられることを前提として、グループ内での細かな違いを無視することになります。

一方の回帰は、データの傾向を見ることが目的です。分類とは反対に、入力されるデータを1つのグループとして扱ったうえで、そのグループ内での違いを分析することになります。

グラフ上にデータがプロット（書き込むこと）されているとすれば、分類はデータ全体をできるだけ分けるように線を引くこと、回帰はデータ全体にできるだけ重なるように線を引くことに相当します。以上はごくざっくりとした整理ですが、まずはそのように捉えてください。

■ 分類と回帰

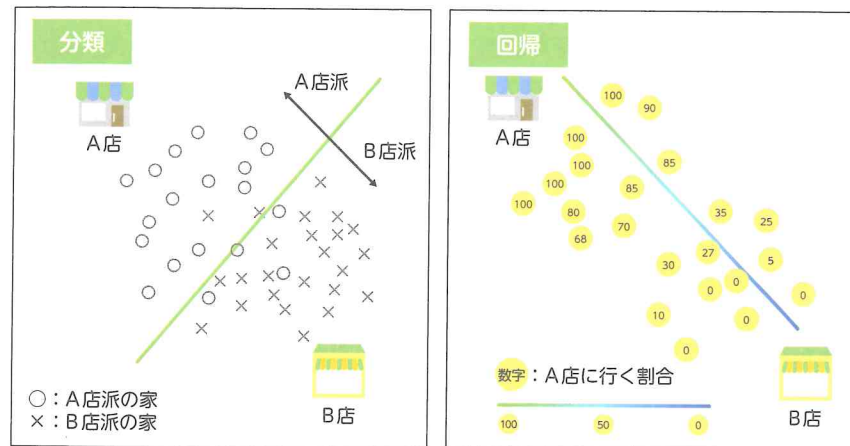


次ページからは具体例です。地図上にA店とB店というチェーンの小売店が2店舗あるとしましょう。そして、地図内に存在する家庭をランダムに抽出し、「A店とB店のどちらをどれくらい利用するか」という割合を調査したとします。そのうえで、まずは分類を行ってみましょう。

分類では、2つの小売店のうち、利用する割合によって、各家庭を「A店派の家庭」と「B店派の家庭」というように分けます。下図左で確認すると、どの家庭も基本的には自宅から近いスーパーを使っていることがわかります。これらを分類するため、地図上に線を書いてみましょう。この線は、A店派とB店派をできるだけ引き離すように書きます。この線を参照すれば、調査を行わなかった家庭であっても、A店派なのかB店派なのか予測が立てやすくなります。

次に、回帰です。回帰では、各家庭におけるA店とB店を利用する割合をそのまま使います。これを地図上にプロットすると、下図右の通り、やはり自宅に近いスーパーを利用する割合が高くなっています。回帰では、この傾向をよく反映するように線を引きます。今回の場合であればA店とB店を結ぶ線を引き、A店側を「A100%」、B店側を「A0% (= B100%)」と定義すればよいのです。この線のどこに近いかで、調査を行わなかった家庭であっても、A店とB店の利用割合を予測できるというわけです。

■ 小売店を例にみる分類と回帰



まとめ

□ 機械学習の問題は大きく分類、回帰に分けられる