

AI 自習ドリル

第16回 回帰問題を利用したデータの補完と予測

足立 悠

前回(2022年4月号)に引き続き、今回もディープ・ラーニングを利用した回帰問題について考えてみましょう。前は入力した特徴量に対し1つの数値を出力

する問題を解きましたが、今回は入力した特徴量に対し複数の数値を出力する問題を解いてみましょう。

1 回帰問題のおさらい

● 回帰問題と分類問題の違い

機械学習の種類は大きく分けて以下の3つがあります。

- 教師あり学習：特徴量と目的変数から状態を予測する
- 教師なし学習：特徴量から状態を要約する
- 強化学習：報酬を獲得する行動を最適化する

ここで教師あり学習は、分類と回帰問題を解くことができます。図1-1のように分類問題では、●と×のクラスを持つ既存データがあるとき、新規データ★

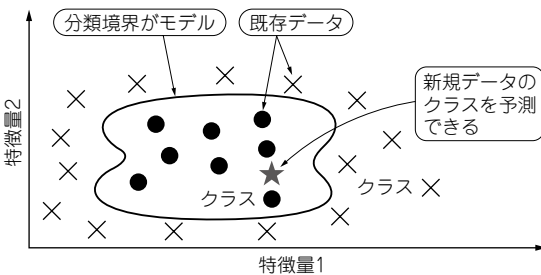


図1-1 分類問題は新規データのクラスを予測する

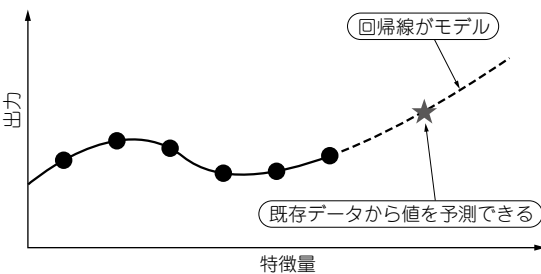


図1-2 回帰問題は過去データのふるまいから値を予測する

のクラスを予測できます。一方、回帰問題では、図1-2のように既存データの挙動から値★を予測できます。それぞれ分類境界と回帰線があり、これらがモデルです。

● 回帰問題を解くアルゴリズム

▶ 単回帰分析…1つの観測値から1つの予測値を得る

まずは、単回帰分析について考えます。これは例えば、身長から体重を予測する、身長から服の大きさを予測するなどが考えられます。観測値と予測値の関係は、次のように表現できます。また、イメージとしては図1-3のようになります。

$$Y=aX+b$$

ここで、 Y は予測値、 X は観測値、 a は回帰係数(傾き)、 b は切片です。この式の右辺 $aX+b$ は、回帰線(回帰モデル)を意味します。 a と b の値は過去のデータ(のパターン)から決めます。この過程を「学習」と呼ぶ

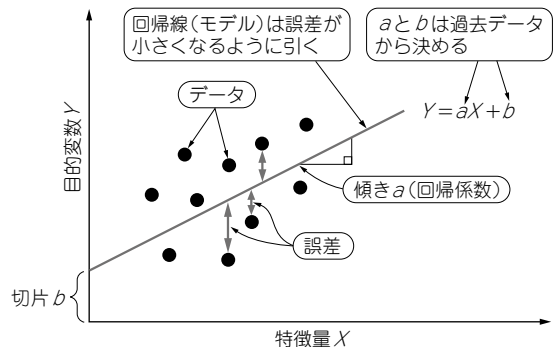


図1-3 単回帰分析では各データとの誤差が最小になる回帰線を引く