第 3 章

線形回帰 (Linear Regression)

この章では、教師つき学習の単純なアプローチの一つである線形回帰について述べる。線形回帰は単純ではあるが、より高度な機械学習を学ぶ前にマスターしておかなければならない事をいくつも含んでいる。

2 章で紹介した広告と売上データを思い出してほしい。ここでは、以下のような問いに答えられることを目指す。

1. 広告出稿と売上に関係があるか？
2. どの程度広告出稿と売上に関係があるのか？
3. 売上に寄与しているのはどのメディアか？
4. 売上における広告出稿それぞれの効果をどれくらい正確に推定できるのか？
5. どの程度正確に未来の売上を予測できるか？
6. 関係性は線形であるか？
7. 広告メディアのシナジーはあるのか？

3.1 単変数線形回帰 (Simple Linear Regression)

連続値 Y を 1 つの説明変数 X で予測することを考える。X と Y に線形関係が近似的にあると仮定する、すなわち、

$$Y \approx \beta_0 + \beta_1 X. \tag{3.1}$$

ここで、”≈”は”近似的にモデル化される”という意味である。よく Y を X で回帰するという。例えば、売上を TV で線形回帰すると、

$$\text{sales} \approx \beta_0 + \beta_1 \times TV.$$

ここで β_0 は切片、β_1 は傾きと呼ぶ。これらを係数といったり、パラメータといったりする。トレーニングデータを用いて、$\hat{\beta}_0$ と $\hat{\beta}_1$ を算出すると、これらを用いて TV ad を説明変数とする予測ができる:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \tag{3.2}$$

ここで、$\hat{y}$ は $X = x$ としたときの予測値である。

3.1.1 係数の算出

係数 β_0、β_1 は未知なので、n 個のデータ

$$(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$$

から算出する必要がある。その際、（トレーニング）データをよく再現できるように $\hat{\beta}_0, \hat{\beta}_1$ を決めるということにする。ここで、”よく再現できる”というのは、データと予測値がどれくらい近いか、もしくは離れているかを判断する、”近さ”を定義する必要がある。よく用いられるのは、*least squares* を最小にするように係数を算出する方法がある。それに代わる方法は 6 章で考える。

データの i 番目に基づく予測を、$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ とする。このとき、$e_i = y_i - \hat{y}_i$ を i 番目の残差 (*residual*) とよぶ。*RSS (residual sum of squares)* を以下の式で定義する:

$$\mathrm{RSS} = e_1^2 + e_2^2 + \cdots + e_n^2,$$

もしくは、

$$\mathrm{RSS} = (y_1 - \hat{\beta}_0 - \hat{\beta}_1 x_1)^2 + (y_2 - \hat{\beta}_0 - \hat{\beta}_1 x_2)^2 + \cdots + (y_n - \hat{\beta}_0 - \hat{\beta}_1 x_n)^2. \tag{3.3}$$

least squares のアプローチは、RSS を最小にするように $\hat{\beta}_0$、$\hat{\beta}_1$ を決める。最小値をとる係数は、以下の式で計算できる（微分して導出も可能であるが、ここでは省略）:

$$\begin{aligned} \hat{\beta}_1 &= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - x)^2}, \\ \hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x}, \end{aligned} \tag{3.4}$$

ここで、$\bar{y} \equiv \frac{1}{n}\sum_{i=1}^n y_i$、$\bar{x} \equiv \frac{1}{n}\sum_{i=1}^n x_i$ であり、標本平均である。

Figure 3.1 に”Advertising”データで線形回帰させたフィッティングを示している。ここで $\hat{\beta}_0 = 7.03, \hat{\beta}_1 = 0.0475$ である。言い換えれば、TV 広告に $\$1,000$ つぎ込むと、大体 47.5 売上が上がる、ということである。Figure 3.2 に RSS を β_0 と β_1 の関数としてプロットしている。RSS 最小のところが、先ほどの値である。

3.1.2 係数算出の精度評価

(2.1) で X と Y に真の関係 $Y = f(X) + \varepsilon$ があったと仮定しており、ここで f は未知の関数、ε は平均 0 のランダム誤差である。もし、f が近似的に線形関数だとすると、関係性は以下の式で表せる:

$$Y = \beta_0 + \beta_1 X + \varepsilon. \tag{3.5}$$

ここで β_0 は切片、すなわち、$X = 0$ のときの Y の期待値である。β_1 は傾きで、X が 1 単位上がるときの平均の増加分を示す。誤差項はこの単純な回帰で捉えられない部分のすべての誤差を含む。