

1 ロジスティック回帰とは

ロジスティック分析は、ある事件（event）が発生するかしないかを直接予測することではなく、その事件が発生する確率を予測する。従って、被説明変数の値は 0 と 1 の間の数値になる。分析結果、被説明変数の値、すなわち確率が 0.5 より大きいとその事件が発生すると予測し、0.5 を下回るとその事件が発生しないと予測する。

つまり、ロジスティック回帰は、ある事象が発生する確率を予測するための手法である。または、クラスに分類分けするためのものである。

ロジスティック回帰を行うときは、説明変数が年収などの量的データで、被説明変数が「ある事象が発生するかしないか」のような質的データ（ex. 車持ってるか）であることが多い。

このように被説明変数が質的データであっても分析ができるよう一般線形モデルを拡張したのが一般化線形モデル（GLM：Generalized Linear Model）である。

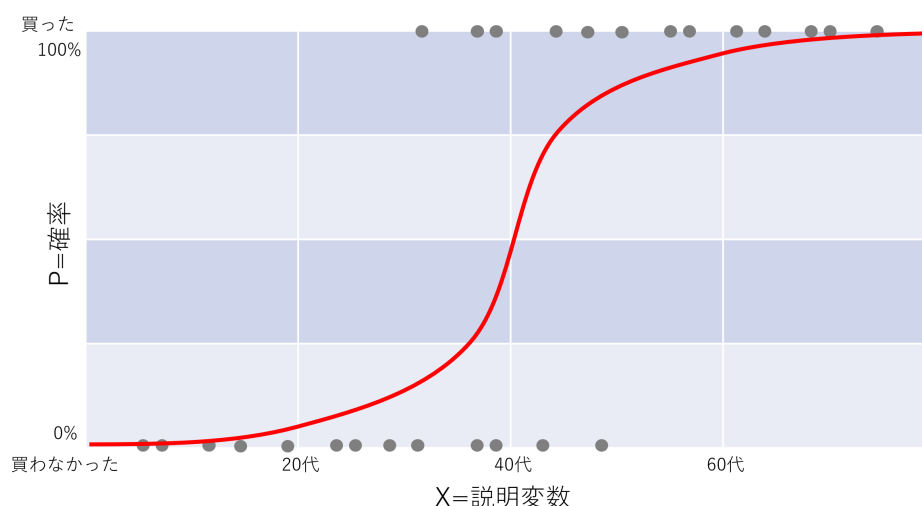


図1 ロジスティック回帰の例

2 ロジスティック回帰の注意点

ロジスティック回帰で解析するべきものを線形回帰で解析すると、誤った結果が出てしまう。つまり、0-1 の間の値をとるべきなのに、それを超えて 0 未満や 1 より大きい値をとってしまう。

3 オッズとオッズ比

オッズとは、ある事象が起こる確率と起こらない確率の比のことである。

$$\text{オッズ} = \frac{\text{起こる確率}}{\text{起こらない確率}} = \frac{p}{1-p} \quad (1)$$

オッズの値は、0 から無限大までの値をとることに注意する。

一方、オッズ比は、ある事象が起こる確率のオッズと、別の事象が起こる確率のオッズの比のことである。

$$\text{オッズ比} = \frac{\frac{p_1}{1-p_1}}{\frac{p_2}{1-p_2}} \quad (2)$$

例えばオッズ比を利用すると、自動車を所有している世帯主が自動車を所有していない世帯主に比べて何倍住宅を所有しているかが計算できる。

ここで、注意しないといけないのが、オッズ比が1 より小さいときである。例えば、「0.01 倍増加する」という結果が得られたとき、これは減少しているということであることに気をつけたい。

4 ロジスティック回帰を使うとき

ロジスティック回帰に適しているのは、求めたい目的変数が明確なシーンである。購入の有無、台風・落雷が発生するかもしれないか、病気に罹るか罹らないかなどの予測にはロジスティック回帰が向いている。逆に目的変数が曖昧なときや、数値を知りたいときにはロジスティック回帰は向かない。

5 ロジット関数

ロジスティック回帰は、線形回帰の結果を 0-1 の間に収めるために、ロジット関数を用いる。

$$\text{ロジット関数} = \text{logit}(p) = \log \frac{p}{1-p} \quad (3)$$

ロジット関数は、オッズの対数をとったものである。

6 ロジスティック回帰の式

回帰において、連続的な値に対しては $ax+b$ というのがもとまる。しかし、今回のように離散的な値を取るにはこれを変える必要がある。そこで登場するのがリンク関数で、ロジスティック回帰には、ロジット関数を使う。(一般化回帰には必須?) (リンク関数によって、連続的な値を離散的な値に変換することができるってことだと思う)

$$\log \frac{p}{1-p} = ax + b \quad (4)$$

$$\Leftrightarrow p = \frac{1}{1 + e^{-(ax+b)}} \leftarrow \text{図1みたいな式} \quad (5)$$

Python で計算した後に出てくる切片と傾きは、この式の a と b だと思われる。

7 評価

分類系のものの評価は、Confusion Matrix を用いるのが良いそう。（大森先生より）

8 ロジスティック回帰の種類

8.1 二項ロジスティック回帰

二項ロジスティック回帰は、被説明変数が 2 値の場合に用いられる。

8.2 多項ロジスティック回帰

多項ロジスティック回帰は、被説明変数が 3 値以上の場合に用いられる。

8.3 順序ロジスティック回帰

順序ロジスティック回帰は、被説明変数が 3 値以上の場合に用いられるが、それらに順位がある場合に用いられる。